Cycles, Contagion and Crises

Papers by:
Tobias Adrian, Hyun Song Shin
Gara Minguez-Afonso
Lev Ratnovski
Prasanna Gai, Sujit Kapadia, Stephen Millard, Ander Perez
Romain Baeriswyl, Camille Cornand
Marcus Miller, Lei Zhang
Sarquis Sarquis
Amil Dasgupta, Roberto Leon Gonzalez, Anja Shortland
Ossip Hühnerbein
Gregor Irwin, Adrian Penalver, Chris Salmon, Ashley Taylor
Patrick Honohan
Nikolaj Schmidt
Arnab Bhattacharjee, Jagjit Chadha, Qi Sun

Editor: Ashley Taylor

SPECIAL PAPER 183

FINANCIAL MARKETS GROUP
SPECIAL PAPER SERIES

Financial Markets Group
RESEARCH CENTRE

This Special Paper is a collection of contributions to a conference on **Cycles, Contagion and Crises** which took place on 28-29 June 2007 at the Financial Markets Group of the London School of Economics and Political Science. The event formed part of the FMG research project 'The Stability of the Global Financial Systems: Regulation and Policy Response' (RES-165-25-0026) funded by the Economic and Social Research Council programme 'World Economy and Finance', Phase I.

We would like to thank the Economic and Social Research Council for their support in making this research possible. For more information on the FMG's activities please visit our website: **http://fmg.lse.ac.uk**

The Financial Markets Group
The London School of Economics and Political Science

August 2008

Cycles, Contagion and Crises

CONTENTS
(conference presenters in bold)

List of Contributors

Jagjit Chadha is Professor of Economics at Kent University. Previously he worked on monetary policy at the Bank of England and was Lecturer at University of Cambridge and Fellow of Clare College, Cambridge. He was Professor of Economics and Chair in Macroeconomics, at University of St Andrews from 2003 and established, as Director, the Centre for Dynamic Macroeconomic Analysis. He joined BNP Paribas as Economic Adviser in 2005, Visiting Professorial Fellow at Centre for International Macroeconomics and Finance, Cambridge University, and member of European Monetary Forum and Euro-Area Business Cycle Network. His current research interests are yield curve modelling and policy under incomplete markets. Current papers are available from Econpapers at **http://econpapers.repec.org/RAS/pch64.htm**.

Camille Cornand holds a PhD from Lyon 2 University (2005, France). She was lecturer at Lyon 2 University from 2002 to 2005. She was post-doc at FMG (LSE) in 2005-2006, financed by the ESRC 'World Economy and Finance Research Programme'. She is now research fellow at CNRS (Centre National de la Recherche Scientifique), affiliated to the BETA (Bureau d'Economie Théorique et Appliquée) at Strasbourg Louis Pasteur University. Her research focuses on the informational aspects of coordination problems in monetary macroeconomics. On the one hand, she has studied the issue of information disclosures and transparency during speculative episodes. On the other hand, her recent research deals with central bank transparency in the conduct of monetary policy. Parts of her work has been published in some journals as The Economic Journal and The Review of International Economics.

Patrick Honohan is Professor of International Financial Economics and Development at Trinity College, Dublin, a position to which he has recently moved from being a Senior Advisor in the World Bank working on issues of financial policy reform. Previously he was Economic Advisor to the Taoiseach and spent several years at the Economic and Social Research Institute, Dublin, and at the Central Bank of Ireland. Patrick Honohan received his PhD from the London School of Economics and has published widely on issues ranging from exchange rate regimes and purchasing-power parity, to microfinance, migration, cost-benefit analysis and statistical methodology. His latest book (with T. Beck) is Making Finance Work for Africa.

Ossip Hühnerbein is a postdoctoral researcher at the Munich Graduate School of Economics where he also undertook his postgraduate studies. He has been working as a research assistant in the World Economy and Finance Programme at the Financial Markets Group at LSE. His research interests include International Finance and Macroeconomics, Sovereign Debt Crises and Contract Theory.

Marcus Miller is Professor of Economics and Associate-Director of ESRC Centre for the Study of Globalisation and Regionalisation at the University of Warwick. Currently holder of a three-year ESRC Professorial Fellowship to study debt, development and the global economy, he is Research Fellow at CEPR, London, visiting Fellow at IIE, Washington, and gives lectures on the Global Economic System at LSE. Educated at Oxford University (PPE) and Yale University (PhD), his previous academic career includes posts at the London School of Economics and Manchester University, with visiting positions at Chicago University Business School and Princeton University. He has worked as an Economist at the Bank of England (and later held a Houblon-Norman Fellowship there); acted as advisor to the Treasury Committee of the House of Commons; was Member and Chair of Academic Panel of the Treasury and was joint director of International Macroeconomics Programme at CEPR from1986 to 1991. Over the years he has been visiting fellow/economic consultant at the OECD, TACIS (Technical Assistance to the CIS, on Macroeconomic Policy in the Ukraine), IMF, World Bank, ECB; and recently at the Inter-American Development Bank.

Gara Mínguez-Afonso obtained a BSc in Business Administration (Finance) from Universidad Carlos III de Madrid in 2003 and an MSc Finance and Economics from LSE in 2004. In September 2004, she began her PhD in Finance at LSE. Gara has been awarded the Bank of Spain scholarship for postgraduate studies for the academic years 2003-04 and 2004-05, the Foundation Caja Madrid scholarship for 2005-06 and the Department of Economics fellowship at Princeton University (2006-07). She joined the FMG in 2005-08. Her research interests centre on financial economics and she is mainly interested in the study of financial crises and market mispricings with the potential to induce destabilising effects on the economy as a whole. Her first paper (Imperfect Common Knowledge in First Generation

* Biographies correct as at time of conference

Models of Currency Crises) has been recently published at the International Journal of Central Banking. Currently she is an Economist at the Federal Reserve Bank of New York.

Ander Perez is a graduate PhD Candidate in the Financial Markets Group and the Economics Department at LSE, with research interests in credit frictions and business cycle fluctuations. Prior to beginning his PhD in 2003, he worked as an Analyst in the investment banking division of Goldman Sachs focusing on European IPOs, secondary equity issues and M&A transactions. His other professional activities include internships in the Bank of England, where he took part in a research project analysing the effect of financial innovation on the likelihood and potential scale of crises, and the International Monetary Fund, where he focused on self-insurance and foreign reserve pooling in emerging market economies. He has a degree in Finance and Economics from Deusto University, in Spain. Currently he is an Assistant Professor at Pompeu Fabra University.

Lev Ratnovski is a research economist at the Bank of England and a PhD candidate at the University of Amsterdam. Lev's work focuses on corporate finance and banking. He is particularly interested in understanding how financial innovation affects financial system performance and stability.

Sarquis J B Sarquis holds a position at the Graduate School of Diplomacy, Brazil and is a PhD student at LSE's Economics Department and member of the FMG. He has worked as an economic adviser at the Brazilian Foreign Service, starting his diplomatic career at the Economic Department of the Ministry of Foreign Affairs (1991-1996), and has worked at the Embassies in London and Berlin as head of the Economic and Financial Section. He holds an MA in Economics, University of Brasília (UnB); and an MSc in Economics, London School of Economics (LSE). His research interests are International Macroeconomics, International Finance, Business Cycles and Credit Cycles. In particular he is interested in business cycles and growth in emerging markets and their interaction with financial constraints.

Nikolaj Schmidt is a PhD student at the Accounting and Finance department at the LSE and a research assistant in the FMG. He holds an MSc in Economics from the University of Copenhagen, and an MSc in Finance and Economics from the LSE. Prior to enrolling in the PhD programme, he worked as a derivatives trader and structurer at Goldman Sachs. His research is on the impact of financial sector liberalisation, with specific focus on the importance of foreign bank entry to developing economies.

Hyun Song Shin is Professor of Economics at Princeton University, affiliated with the Department of Economics and the Bendheim Center for Finance, and is co-director of the Financial Regulation programme at the FMG. Prior to leaving for Princeton last year, he was Professor of Finance at the London School of Economics. His research interests are in financial economics and economic theory with particular reference to financial crises, disclosures, risk and financial stability issues, topics on which he has published widely both in academic and practitioner outlets. He has served as editor or editorial board member of several scholarly journals, and has served in an advisory capacity to central banks and policy organisations on financial stability issues. He is a fellow of the Econometric Society and of the British Academy.

Anja Shortland is a lecturer in Economics at Brunel University. She holds an MEng from Oxford University and an MSc and PhD in International Relations from the London School of Economics. Her previous position was at Leicester University (2001-2006) and she has also taught at LSE, Oxford and SOAS. Her research interests lie in the interaction between political conditions, institutions and economic outcomes. She has published on the economic effects of political violence, the politics of financial development on government ownership of banks with papers forthcoming in the Journal of Development Economics and Oxford Economic Papers.

Ashley Taylor is a graduate Economics PhD student at LSE and research assistant in the FMG with research interests in international finance and macroeconomics and issues related to financial stability and regulation. His PhD thesis research focused on the macroeconomic implications of increasing international financial integration. Following completion of undergraduate studies at Oxford University in 1997 he worked as an economic consultant before taking the Economics MSc at LSE in 2000. He then joined the Bank of England, working on financial stability issues, prior to starting his PhD in 2003. He is currently working at the Young Professional's Programme at the World Bank.

Foreward

On 28-29 June 2007, the Financial Markets Group (FMG) of the London School of Economics and Political Science (LSE) organised a conference on ‘Cycles, Contagion and Crises’. The event formed part of a research project on ‘The Stability of the Global Financial Systems: Regulation and Policy Response’ funded by the Economic and Social Research Council (ESRC) Research Programme on ‘World Economy and Finance’. In this volume the FMG/LSE publishes a selection of papers presented at the conference. We would like to thank all conference participants for their contribution to the day’s proceedings and this publication. Finally, we are indebted to the FMG/LSE’s administrative staff, who provided invaluable assistance in the organisation of the conference.

Organised by Professors Charles Goodhart (FMG/LSE) and Hyun Shin (FMG/LSE and Princeton), the objective of the event was to report research in this field by established academics alongside that of FMG junior researchers affiliated with the project who received valuable comments from an audience of senior researchers and practitioners. The Conference covered topics under all three themes of its title, ‘Cycles, Contagion and Crises’, from the perspective of both developed and emerging economies. The set of papers in this volume are organised into the following broad areas: financial market liquidity and innovation; transparency and monetary policy; real business cycle models; crises in emerging markets, their contagion and policy response; and, banking in emerging and developing economies.

Financial Market Liquidity and Innovation

The conference papers presented on financial market liquidity and innovation have proved to be extremely prescient and timely in light of the ongoing credit crisis which began over the summer of 2007. For example, Hyun Song Shin (Princeton University) presented co-authored work with Tobias Adrian (Federal Reserve Bank of New York) on ‘Liquidity and Leverage’.* The focus of the paper is the management of leverage by financial intermediaries and the resulting aggregate implications for liquidity and risk appetite. The leverage of major financial intermediaries who mark their assets to market would be expected to be countercyclical if there was no balance sheet adjustment when their net worth changes due to asset price movements. Thus, the evidence of procyclical leverage ratios presented in the paper implies active balance sheet management as such institutions manage their value-at-risk in line with models of risk and economic capital. Although such decisions may be optimal from the point of view of an individual institution they do not take into account the aggregate consequences for liquidity and asset prices. For example, it is shown in the paper that expansions and contractions of balance sheets can forecast shifts in risk appetite measured by the difference between the VIX index and realised volatility. The paper then goes on to discuss a notion of aggregate liquidity in financial markets as determined by the rate of growth of aggregate balance sheets. When balance sheets are generally strong leverage may be too low and so banks expand their assets, for example through searching for additional borrowers such as sub-prime mortgage borrowers. In doing so, credit quality may fall which, as we have seen in recent months, may be a driver of future financial crises.

Financial market liquidity was also the topic of the paper by Gara Minguez-Afonso (Princeton University and FMG/LSE), ‘Liquidity and Congestion’. This work focuses on the impact of new investors on liquidity within a search-based model of asset trading. In such a set-up the entry of new investors not only reduces search costs (attracting potential investors) but may lead to ‘congestion effects’ as investors concentrate on one side of the market. The trade-off between these two effects determines the equilibrium level of market liquidity. If the congestion effect dominates then the model predicts some interesting effects. For example, reducing market frictions could even reduce liquidity and welfare.

The paper by Lev Ratnovski (Bank of England and University of Amsterdam) also focused on liquidity but from the perspective of the risk management of individual banks. ‘Liquidity and Transparency in Bank Risk Management’ analyses the sources of financial instability that arise from the mismanagement of liquidity and transparency carried out by commercial banks. It is claimed that both

* The former title of the paper, as presented at the conference, was ‘Liquidity and Financial Cycles’.

are important components of risk management, as liquidity buffers provide complete insurance against small liquidity needs, and transparency offers partial insurance against large ones. The paper notes that, due to leverage, banks may under invest in both types of insurance, and that while liquidity buffers can be imposed, an adequate level of transparency may be more difficult to regulate.

In another timely analysis, Ander Perez (FMG/LSE) presented joint work with Prasanna Gai, Sujit Kapadia, Stephen Millard (Bank of England) on 'Financial Innovation, Macroeconomic Stability and Systemic Crises'. The paper studies the effect of two trends over recent decades, namely the decrease in macroeconomic volatility and the ongoing process of rapid financial innovation, on the likelihood and potential scale of systemic crises. The paper finds that these changes lead to a financial system that is more resilient to moderate shocks and hence that crises are less likely, but that should a sufficiently severe shock occur, the severity of the crisis could be significantly larger.

Transparency and Monetary Policy

Camille Cornand (BETA/CNRS – Université Louis Pasteur Strasbourg) presented joint work with Romain Baeriswyl (Ludwig-Maximilians Universität München) on 'Monetary Policy and its Informative Value'. Recognising the dual role of monetary policy as both a stabilising central bank instrument and a signal about the state of the economy, the paper analyses the optimal monetary policy and the optimal disclosure strategy. It is argued that greater transparency is desirable when the level of complementarities between an individual firm's prices and that of others is low, when supply shocks are not too volatile, when the central bank is more inclined towards stabilising prices than output, and when firms have relatively precise information about the economy.

Emerging Markets Crises

A variety of economic and policy issues arising from emerging market crises were examined during the conference. These ranged from the impact of crises on growth, their transmission to other countries, through to policy topics such as the promotion of collective action clauses in bonds and the determinants of IMF subscription levels.

In 'Crises and recovery in emerging markets: 'Phoenix miracles' or endogenous growth?', Marcus Miller (Warwick University) and co-author Lei Zhang (University of Warwick) examine the impact of currency crises and the subsequent output recovery in emerging economies. The starting point for the analysis is the view put forward by Calvo and co-authors at the Inter-American Development Bank of crises as supply-side "phoenix miracles" whereby factor productivity falls in recession, but rises promptly thereafter. Instead, with reference to the East Asian crises, the authors argue that external shocks and the resulting balance sheet effects interrupted economic growth which resumed after the recession, albeit at a lower level of GDP. The authors use an endogenous growth model of the supply-side to which they add the balance sheet effects emphasised in Aghion, Bacchetta and Banerjee (2000).*

The paper presented by Sarquis Sarquis (FMG/LSE and Graduate School of Diplomacy, Brazil) analyses 'Interest Rate and Business Cycles in a Credit Constrained Small Open Economy'. The motivation for the subsequent theoretical analysis is provided by VAR analysis of Brazil's business cycle. This empirical work suggests that shocks to the exogenous external real interest rate faced by the economy drive much of business cycle volatility. The paper then presents a standard business cycle small open economy model with the addition of an endogenous collateral constraint on foreign liabilities. The proposed model reveals considerable propagation of shocks, particularly in relation to interest rate innovations, and appears to match most of the empirical regularities presented.

In 'Regionality Revisited: An Examination of the Direction of Spread of Currency Crises', Anja Shortland (Brunel University) joint with Amil Dasgupta (FMG/LSE) and Roberto Leon Gonzalez

* Phlilppe Aghion, Phlippe Bacchetta and Abhijit Banerjee (2000), A Simple Model of Monetary Policy and Currency Crises', *European Economic Review* 44 (4-6), 728-738.

(University of Leicester) examine regionality and the directions of spread of currency crises. Using Bayesian methodologies, they examine currency crises in the nineties and find an important role for institutional similarity to the ground-zero country (as measured by quality-of-governance indicators) in determining the direction of contagion in emerging market currency crises. Trade competition and financial links are other drivers, but prove rather sensitive to periods and priors. Therefore, the authors favour the 'wake up call' hypothesis for financial contagion.

One of the policy responses to the emerging economy crises of the 1990s was the promotion of reforms to sovereign bond contracts such as the introduction of collection action clauses (CACs). These clauses were designed to overcome creditor coordination problems through the inclusion of certain provisions, for example that the terms of the bond contract can be changed by a qualified majority of outstanding debt holders. In 'Coordination Clauses and the Price of Sovereign Debt', Ossip Huhnerbein (Munich Graduate School of Economics) first develops a theoretical model to consider the impact of such clauses on the pricing on sovereign debt. Then, the model's implications are tested using a spread data for a sample of bonds issued by 19 emerging economies maturing after 2003. The regressions suggest that CACs have positive direct effects on bond spreads in secondary markets but negative interaction effects with the share of bonds with CACs. The former is interpreted as revealing the seniority of bonds without CACs.

In 'Dealing with country diversity: challenges for the IMF credit union model', Ashley Taylor (FMG/LSE) and co-authors Gregor Irwin, Adrian Penalver and Chris Salmon (Bank of England) examine the implications of changes in the characteristics of IMF members, in particular their likelihood of crises, for the political determination of subscription levels. A simple theoretical model of the IMF as a credit union is developed where the membership decides on the Fund size and hence the amount of crisis lending it provides. The model is used to analyse equilibrium Fund subscriptions and country reserve choices under three different characterisations of the Fund's decision-making processes: unconstrained majority voting, constrained majority voting, and qualified majority voting with an agenda setter. The paper concludes with a discussion of how the Fund's financial structure has become less well-suited for a world in which member countries differ sharply in their economic characteristics and needs.

Banking in Emerging and Developing Economies

Banking in emerging and developing economies were covered in two papers presented at the conference. Patrick Honohan (Trinity College Dublin) focused on the appropriate regulatory approach which should be taken in considering 'Should bank supervisors in developing countries exercise more or less forbearance?'. Of particular importance is the tradeoff between regulatory forbearance potentially deepening banking crises versus the potential that a lack of forbearance could lead to bank closures thus precipitating a crisis. Whilst in theory forbearance may yield benefits in more sophisticated regulatory environments the paper considers whether this applies to developing economies. Three key differences in developing economies are emphasised: namely worse information on banks' balance sheets; less interdependence between banks; and greater agency problems affecting regulators. Although empirical evidence on the effectiveness or risks of policies of forbearance are difficult to obtain the above-mentioned features of their financial systems suggest that there should be less forbearance in developing economies.

In 'Foreign Bank Entry: A Liquidity Based Theory of Entry and Credit Market Segmentation', Nikolaj Schmidt (FMG/LSE) focused on the implications of foreign bank entry in emerging and developing economies. In particular, the paper develops a theoretical model to analyse how local credit markets in such economies are affected by the entry of international financial intermediaries. The paper stresses that the entry of these international banks allows the provision of funding to solvent (exporting) firms during liquidity shortages in the local financial system. Features of the borrowers' business, rather than information about the borrower, then become the driver of the clientele effects following foreign bank entry. However, foreign bank entry may increase the vulnerability to liquidity shocks of the domestic financial sector.

Real business cycle models

A number of the papers in the conference used different versions of real business cycle models to examine cyclical trends in developed economies and the interaction between the financial and real sectors. For example, the paper by Jagjit Chadha (University of Kent), joint with Arnab Bhattacharjee (University of St. Andrews) and Qi Sun (University of St. Andrews), examines the question of 'Productivity, Preferences and UIP deviations in an Open Economy Business Cycle Model'.* The work aims to account for two important puzzles in international business cycle theory: the absence of complete international risk sharing across open economies and the related disconnect between the real exchange rate and relative consumption. In doing so it employs a two-country, two-sector, flexible price dynamic stochastic general equilibrium model. The model embodies shock processes to productivity in the tradable and non-tradable sectors, to consumer preferences between leisure and work and to deviations in the exchange rate from uncovered interest parity. It is found that a combination of these shock processes provides some movement towards resolution of the above puzzles. In doing so the paper introduces a suite of econometric tests to compare the model's fit to the observed data.

Ashley Taylor
August 2008

* The former title of the paper, as presented at the conference, was 'Can Open Economy Business Cycle Models Explain Business Cycle Facts?'.

Liquidity and Leverage*

Tobias Adrian
Federal Reserve Bank of New York
tobias.adrian@ny.frb.org

Hyun Song Shin
Princeton University
hsshin@princeton.edu

September 2007

Abstract

In a financial system where balance sheets are continuously marked to market, asset price changes show up immediately in changes in net worth, and elicit responses from financial intermediaries who adjust the size of their balance sheets. We document evidence that marked-to-market leverage is strongly procyclical. Such behavior has aggregate consequences. Changes in aggregate balance sheets for intermediaries forecast changes in risk appetite in financial markets, as measured by the innovations in the VIX index. Aggregate liquidity can be seen as the rate of change of the aggregate balance sheet of the financial intermediaries.

*A previous version of this paper was presented at the 6th BIS Annual Conference, "Financial System and Macroeconomic Resilience", 18-19 June 2007 under its former title "Liquidity and Financial Cycles". We thank conference participants at the BIS conference, and seminar participants at the Federal Reserve Bank of New York, the Federal Reserve Bank of Chicago, and Princeton University for their comments. The views expressed in this paper are those of the authors and do not necessarily represent those of the Federal Reserve Bank of New York or the Federal Reserve System.

1. Introduction

In a financial system where balance sheets are continuously marked to market, changes in asset prices show up immediately on the balance sheet, and so have an immediate impact on the net worth of all constituents of the financial system. The net worth of financial intermediaries are especially sensitive to fluctuations in asset prices given the highly leveraged nature of such intermediaries' balance sheets.

Our focus in this paper is on the reactions of the financial intermediaries to changes in their net worth, and the market-wide consequences of such reactions. If the financial intermediaries were passive and do not adjust their balance sheets to changes in net worth, then leverage would fall when total assets rise. Change in leverage and change in balance sheet size would then be negatively related.

However, as we will see below, the evidence points to a strongly *positive* relationship between changes in leverage and changes in balance sheet size. Far from being passive, the evidence points to financial intermediaries adjusting their balance sheets actively, and doing so in such a way that leverage is high during booms and low during busts. That is, leverage is procyclical.

Procyclical leverage can be seen as a consequence of the active management of balance sheets by financial intermediaries who respond to changes in prices and measured risk. For financial intermediaries, their models of risk and economic capital dictate active management of their overall value at risk (VaR) through adjustments of their balance sheets.

From the point of view of each financial intermediary, decision rules that result in procyclical leverage are readily understandable. However, there are aggregate consequences of such behavior for the financial system as a whole that are not taken into consideration by an individual financial institution. We exhibit evidence that procyclical leverage has spillover effects at the aggregate level through shifts in risk appetite and funding liquidity. In particular, balance sheet fluctuations forecast shifts in risk appetite, as measured by the VIX index.

Our paper has two main objectives. Our first objective is to document the determinants of balance sheet size and leverage for the group of financial intermediaries (including the major Wall Street investment banks) that operate primarily through the capital markets. We show that leverage is strongly procyclical for these intermediaries, and that the margin of adjustment on the balance sheet is through repos and reverse repos (and other collateralized borrowing and lending). In turn, procyclical leverage can be attributed to the bank's capital allocation decision that rests on measured risks ruling at the time. We find that the value-at-risk (VaR) disclosed by the banks is an important determinant of balance sheet stance, but we also find evidence of an additional procyclical element in leverage that operates over and above that implied by their disclosed value-at-risk.

Our second objective is to pursue the aggregate consequences of such procyclical leverage, and document evidence that expansions and contractions of balance sheets have important asset pricing consequences through shifts in market-wide risk appetite. In particular, we show that changes in aggregate intermediary balance sheet size can forecast innovations in market-wide risk premiums as measured by the VIX index of implied volatility in the stock market. We see this as an important empirical finding. Previous work in asset pricing has shown that innovations in the VIX index capture key components of asset pricing that conventional empirical models have been unable to address fully. By being able to forecast shifts in risk appetite, we hope to inject a new element in thinking about risk appetite and asset prices. The shift in risk appetite is closely related to other notions of liquidity, such as the notion of "funding liquidity" used by

Brunnermeier and Pedersen (2005b)[1]. One of our contributions is to explain the origins of funding liquidity in terms of financial intermediary behavior.

Our findings also shed light on the concept of "liquidity" as used in common discourse about financial market conditions. In the financial press and other market commentary, asset price booms are sometimes attributed to "excess liquidity" in the financial system. Financial commentators are fond of using the associated metaphors, such as the financial markets being "awash with liquidity", or liquidity "sloshing around". However, the precise sense in which "liquidity" is being used in such contexts is often left unspecified.

Our empirical findings suggest that funding liquidity can be understood as the rate of growth of aggregate balance sheets. When financial intermediaries' balance sheets are generally strong, their leverage is too low. The financial intermediaries hold surplus capital, and they will attempt to find ways in which they can employ their surplus capital. In a loose analogy with manufacturing firms, we may see the financial system as having "surplus capacity". For such surplus capacity to be utilized, the intermediaries must expand their balance sheets. On the liabilities side, they take on more short-term debt. On the asset side, they search for potential borrowers that they can lend to. Funding liquidity is intimately tied to how hard the financial intermediaries search for borrowers.

The outline of our paper is as follows. We begin with a review of some very basic balance sheet arithmetic on the relationship between leverage and total assets. The purpose of this initial exercise is to motivate our empirical investigation of the balance sheet changes of financial intermediaries in section 3. Having outlined the facts, in section 4, we show that changes in aggregate repo positions of the major financial intermediaries can forecast innovations in the volatility risk-premium, where the volatility risk premium is defined as the difference between the VIX

[1] See also Gromb and Vayanos (2002).

index and realized volatility. We conclude with discussions of the implications of our findings for funding liquidity.

2. Some Basic Balance Sheet Arithmetic

What is the relationship between *leverage* and *balance sheet size*? We begin with some very elementary balance sheet arithmetic, so as to focus ideas. Before looking at the evidence for financial intermediaries, let us think about the relationship between balance sheet size and leverage for a household. The household owns a house financed with a mortgage. For concreteness, suppose the house is worth 100, the mortgage value is 90, and so the household has net worth (equity) of 10. The initial balance sheet then is given by:

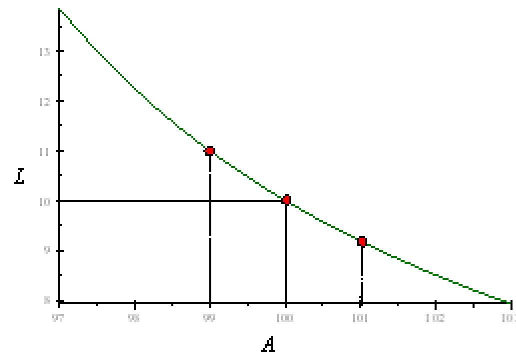
Assets	Liabilities
100	10
	90

Leverage is defined as the ratio of total assets to equity, hence is $100/10 = 10$. What happens to leverage as total assets fluctuate? Denote by A the market value of total assets and E is the market value of equity. We make the simplifying assumption that the market value of debt stays roughly constant at 90 for small shifts in the value of total assets. Total leverage is then

$$L \simeq \frac{A}{A - 90}$$

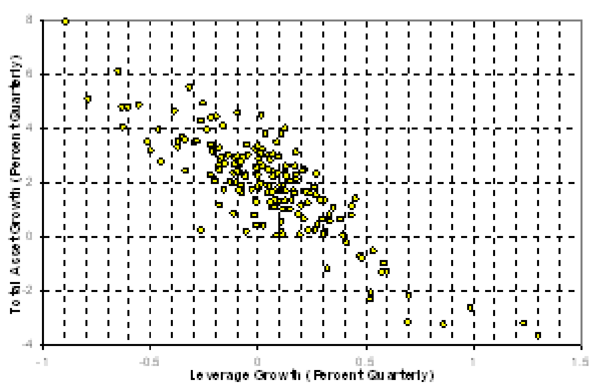
Leverage is inversely related to total assets. When the price of my house goes up, my net worth increases, and so my leverage goes down. Figure 2.1 illustrates the negative relationship between total assets and leverage. Indeed, for households, the negative relationship between total assets and leverage is clearly borne out in the aggregate data. Figure 2.2 plots the quarterly changes in total assets to quarterly changes in leverage as given in the Flow of Funds account for the United

Figure 2.1: Leverage for passive investor



States. The data are from 1963 to 2006. The scatter chart shows a strongly negative relationship, as suggested by Figure 2.1.

Figure 2.2: Total Assets and Leverage of Household.



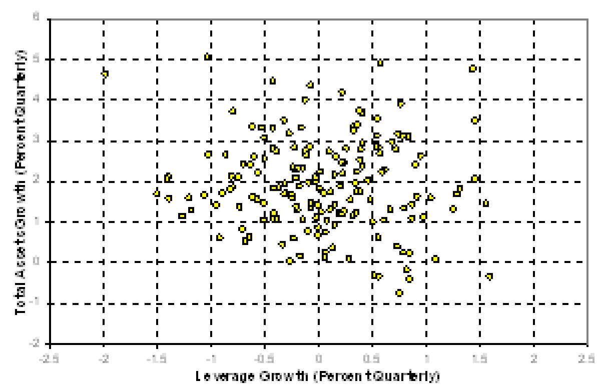
We can ask the same question for firms, and we will address this question for three different types of firms.

- Non-financial firms

- Commercial banks
- Security brokers and dealers (including investment banks).

If a firm were passive in the face of fluctuating asset prices, then leverage would vary inversely with total assets. However, the evidence points to a more active management of balance sheets. Figure 2.3 is a scatter chart of the change in

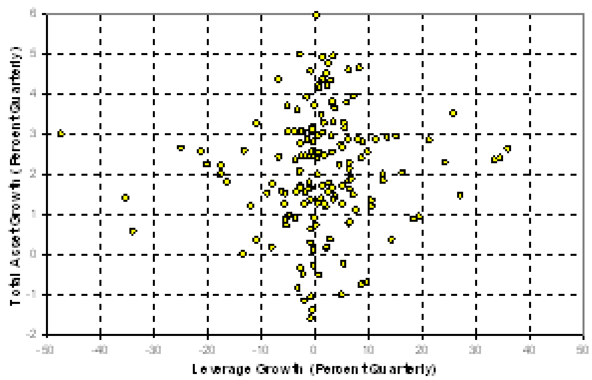
Figure 2.3: Total Assets and Leverage of Non-financial, Non-farm Corporates



leverage and change in total assets of non-financial, non-farm corporations drawn from the U.S. flow of funds data (1963 to 2006). The scatter chart shows much less of a negative pattern, suggesting that companies react to changes in assets by shifting their stance on leverage.

More notable still is the analogous chart for U.S. commercial banks, again drawn from the U.S. Flow of Funds accounts. Figure 2.4 is the scatter chart plotting changes in leverage against changes in total assets for U.S. commercial banks. A large number of the observations line up along the vertical line that passes through zero change in leverage. In other words, the data show the outward signs of commercial banks targeting a fixed leverage ratio.

Figure 2.4: Total Assets and Leverage of Commercial Banks

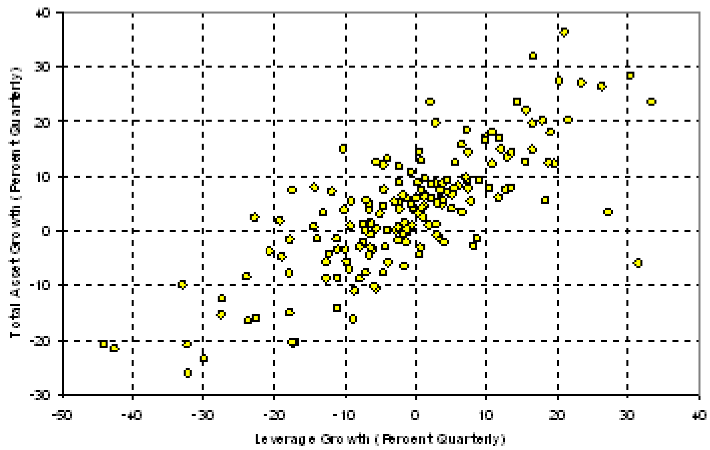


However, even more striking than the scatter chart for commercial banks is that for security dealers and brokers, that include the major Wall Street investment banks. Figure 2.5 is the scatter chart for U.S. security dealers and brokers, again drawn from the Flow of Funds accounts (1963 - 2006). The alignment of the observations is now the reverse of that for households. There is a strongly *positive* relationship between changes in total assets and changes in leverage. In this sense, leverage is pro-cyclical.

In order to appreciate the aggregate consequences of pro-cyclical leverage, let us first consider the behavior of a financial intermediary that manages its balance sheet actively to as to maintain a *constant* leverage ratio of 10. Suppose the initial balance sheet is as follows. The financial intermediary holds 100 worth of securities, and has funded this holding with debt worth 90.

Assets	Liabilities
Securities, 100	Equity, 10
	Debt, 90

Figure 2.5: Total Assets and Leverage of Security Brokers and Dealers



Assume that the price of debt is approximately constant for small changes in total assets. Suppose the price of securities increases by 1% to 101.

Assets	Liabilities
Securities, 101	Equity, 11
	Debt, 90

Leverage then falls to $101/11 = 9.18$. If the bank targets leverage of 10, then it must take on additional debt of D to purchase D worth of securities on the asset side so that

$$\frac{\text{assets}}{\text{equity}} = \frac{101 + D}{11} = 10$$

The solution is $D = 9$. The bank takes on additional debt worth 9, and with this money purchases securities worth 9. Thus, an increase in the price of the security of 1 leads to an increased holding worth 9. The demand curve is *upward-sloping*. After the purchase, leverage is now back up to 10.

Assets	Liabilities
Securities, 110	Equity, 11
	Debt, 99

The mechanism works in reverse, too. Suppose there is shock to the securities price so that the value of security holdings falls to 109. On the liabilities side, it is equity that bears the burden of adjustment, since the value of debt stays approximately constant.

Assets	Liabilities
Securities, 109	Equity, 10
	Debt, 99

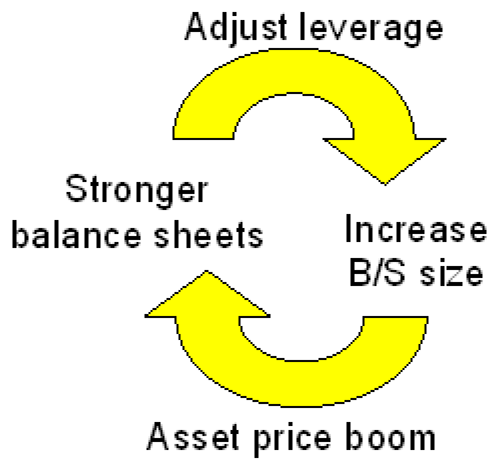
Leverage is now too high ($109/10 = 10.9$). The bank can adjust down its leverage by selling securities worth 9, and paying down 9 worth of debt. Thus, a *fall* in the price of securities of leads to *sales* of securities. The supply curve is *downward*-sloping. The new balance sheet then looks as follows.

Assets	Liabilities
Securities, 100	Equity, 10
	Debt, 90

The balance sheet is now back to where it started before the price changes. Leverage is back down to the target level of 10.

Leverage targeting entails upward-sloping demands and downward-sloping supplies. The perverse nature of the demand and supply curves are even stronger when the leverage of the financial intermediary is pro-cyclical - that is, when leverage is high during booms and low during busts. When the securities price

Figure 2.6: Adjustment of Leverage in Booms

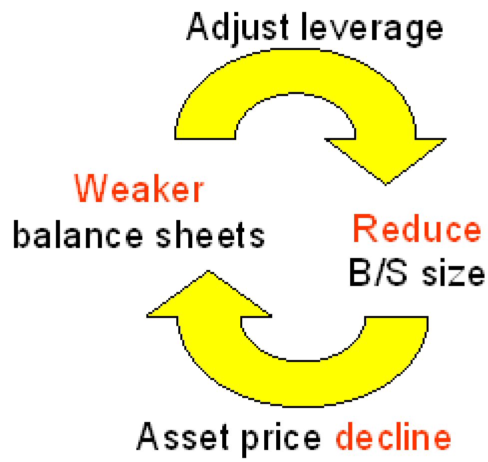


goes up, the upward adjustment of leverage entails purchases of securities that are even larger than that for the case of constant leverage. If, in addition, there is the possibility of feedback, then the adjustment of leverage and price changes will reinforce each other in an amplification of the financial cycle.

If we hypothesize that greater demand for the asset tends to put upward pressure on its price (a plausible hypothesis, it would seem), then there is the potential for a feedback effect in which stronger balance sheets feed greater demand for the asset, which in turn raises the asset's price and lead to stronger balance sheets. Figure 2.6 illustrates the feedback during a boom. The mechanism works exactly in reverse in downturns. If we hypothesize that greater supply of the asset tends to put downward pressure on its price, then there is the potential for a feedback effect in which weaker balance sheets lead to greater sales of the asset, which depresses the asset's price and lead to even weaker balance sheets. Figure 2.7 illustrates the feedback during a downturn.

In section 4, we return to the issue of feedback by exhibiting evidence that is consistent with the amplification effects sketched above. We will see that

Figure 2.7: Leverage Adjustment in Downturn



changes in key balance sheet components forecast changes in the VIX index of implied volatility in the stock market.

3. A First Look at the Evidence

3.1. Investment Bank Balance Sheets

To set the stage for our empirical study, we begin by examining the quarterly changes in the balance sheets of five large investment banks, as listed below in Table 1. The data are drawn from the Mergent database, which in turn are based on the regulatory filings with the U.S. Securities and Exchange Commission (SEC) on their 10-K and 10-Q forms.

Table 1: Investment Banks

Name	Sample
Bear Stearns	1997 Q1 – 2007 Q1
Goldman Sachs	1999 Q2 – 2007 Q1
Lehman Brothers	1993 Q2 – 2007 Q1
Merrill Lynch	1991 Q1 – 2007 Q1
Morgan Stanley	1997 Q2 – 2007 Q1

Our choice of these five banks is motivated by our concern to examine "pure play" investment banks that are not part of a larger commercial banking group so as to focus attention on their behavior with respect to the capital markets[2]. Citigroup reported its investment banking operations separately from its commercial banking operations until 2004 as "Citigroup Global Markets", and we have data for the period 1998Q1 to 2004Q4. In some of our charts below, we will report Citigroup Global Markets for comparison for reference. The stylized balance sheet of an investment bank is as follows.

Assets	Liabilities
Trading assets	Short positions
Reverse repos	Repos
Other assets	Long term debt
	Shareholder equity

On the asset side, traded assets are valued at market prices or are short term collateralized loans (such as reverse repos) for which the discrepancy between face value and market value are very small due to the very short term nature of the loans. On the liabilities side, short positions are at market values, and repos are very short term borrowing. We will return to a more detailed descriptions of repos and reverse repos below. Long-term debt is typically a very small fraction of the balance sheet.[3] For these reasons, investment banks provide a good

[2]Hence, we do not include JP Morgan Chase, Credit Suisse, Deutsche Bank, and other brokerage operations that are part of a larger commercial bank.

[3]The balance sheet of Lehman Brothers as of November 2005 shows that short positions are around a quarter of total assets, and long term debt is an even smaller fraction. Shareholder

approximation of the balance sheet that is continuously marked to market, and hence provide insights into how leverage changes with balance sheet size.

The second reason for our study of investment banks lies in their continuously increasing significance for the financial system.

Figure 3.1:

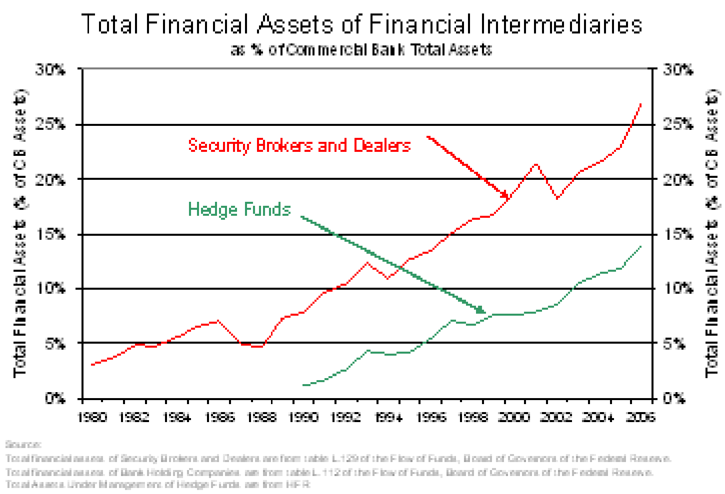


Figure 3.1 plots the size of securities firms' balance sheets relative to that of commercial banks. We also plot the assets under management for hedge funds, although we should be mindful that "assets under management" refers to total shareholder equity, rather than the size of the balance sheet. To obtain total balance sheet size, we should multiply by leverage. Figure 3.1 shows that when expressed as a proportion of commercial banks' balance sheets, securities firms have been increasing their balance sheets at a very rapid rate. Note that when hedge funds' assets under management is converted to balance sheet size by multiplying by a conservative leverage factor of 2, the combined balance sheets

equity is around 4% of total assets (implying leverage of around 25). Short-term borrowing in terms of repurchase agreements and other collateralized borrowing takes up the remainder.

of investment banks and hedge funds is over 50% of commercial banks balance sheets.

Size is not the only issue. When balance sheets are marked to market, the responses to price changes may entail responses that may be disproportionately large. LTCM's balance sheet was small relative to the total financial sector, but its impact would have been underestimated if only size had been taken into account. Similarly, the size of the sub-prime mortgage exposures was small relative to the liabilities of the financial system as a whole, but the credit crisis of 2007 demonstrates that its impact can be large. Table 2 gives the summary statistics of the investment banks over the sample period.

[Table 2]

We begin with the key question left hanging from the previous section. What is the relationship between leverage and total assets? The answer is provided in the scatter charts in figure 3.3. We have included the scatter chart for Citigroup Global Markets (1998Q1 - 2004Q4) for comparison, although Citigroup does not figure in the panel regressions reported below. The scatter chart shows the growth in assets and leverage at a quarterly frequency. In all cases, leverage is large when total assets are large. Leverage is pro-cyclical.

There are some notable common patterns in the scatter charts, but also some notable differences. The events of 1998 are clearly evident in the scatter charts. The early part of the year saw strong growth in total assets, with the attendant increase in leverage. However, the third and fourth quarters of 1998 shows all the hallmarks of financial distress and the attendant retrenchment in the balance sheet. For most banks, there were very large contractions in balance sheet size in 1998Q4, accompanied by large falls in leverage. These points are on the bottom left hand corners of the respective scatter charts, showing large contractions in

Figure 3.2:

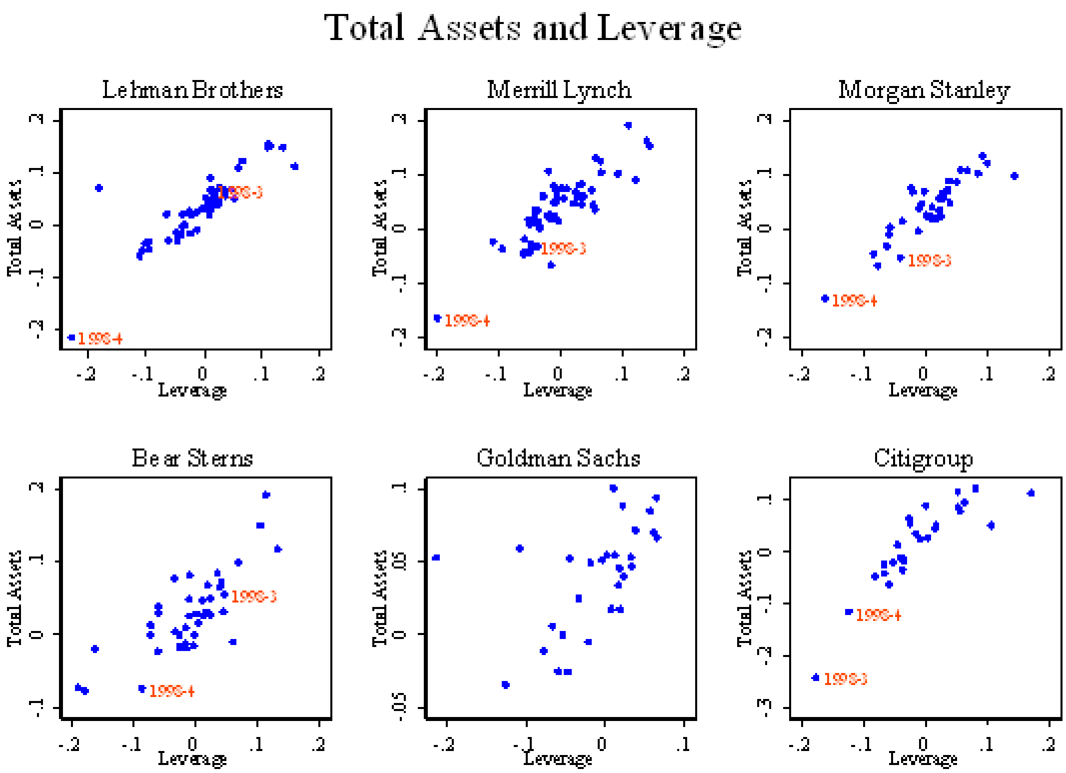


Figure 3.3:

the balance sheet and decrease in leverage. Lehman Brothers and Merrill Lynch seem especially hard hit in 1998Q4.

However, there are also some notable differences. It is notable, for instance, that for Citigroup Global Markets, the large retrenchment seems to have happened in the third quarter of 1998, rather than in the final quarter of 1998. Such a retrenchment would be consistent with the closing down of the former Salomon Brothers fixed income arbitrage desk on July 6th 1998 following the acquisition of the operation by Travelers Group (later, Citigroup). Many commentators see this event as the catalyst for the sequence of events that eventually led to the demise of Long Term Capital Management (LTCM) and the associated financial distress in the summer and early autumn of 1998.[4]

[Table 3]

Table 3 shows the results of a panel regression for change in leverage. The negative relationship between the change in leverage and change in total assets is confirmed in the final column (column (v)) of Table 3. The coefficient on lagged leverage (i.e. previous quarter's leverage) is negative, suggesting that there is mean-reversion in the leverage ratio for the banks. Leverage is positively related to repos.

More interestingly, the regressions reveal which items on the balance sheet are adjusting when balance sheets expand and contract. In particular, the regressions show that the margin of adjustment in the expansion and contraction of balance sheets is through repos. In a repurchase agreement (repo), an institution sells a security while simultaneously agreeing to buy it back at a pre-agreed price on a fixed future date. Such an agreement is tantamount to a collateralized loan, with

[4]The official account (BIS, 1999) is given in the report of the CGFS of the Bank for International Settlements (the so-called "Johnson Report"). Popular accounts, such as Lowenstein (2000) give a description of the background and personalities.

the interest on the loan being the excess of the repurchase price over the sale price. From the perspective of the funds lender – the party who buys the security with the undertaking to re-sell it later – such agreements are called reverse repos. For the buyer, the transaction is equivalent to granting a loan, secured on collateral.

Repos and reverse repos are important financing activities that provide the funds and securities needed by investment banks to take positions in financial markets. For example, a bank taking a long position by buying a security needs to deliver funds to the seller when the security is received on settlement day. If the dealer does not fully finance the security out of its own capital, then it needs to borrow funds. The purchased security is typically used as collateral for the cash borrowing. When the bank sells the security, the sale proceeds can be used to repay the lender.

Reverse repos are loans made by the investment bank against collateral. The bank's prime brokerage business vis-à-vis hedge funds will figure prominently in the reverse repo numbers. The scatter chart gives a glimpse into the way in which changes in leverage are achieved through expansions and contractions in the collateralized borrowing and lending. We saw in our illustrative section on the elementary balance sheet arithmetic that when a bank wishes to expand its balance sheet, it takes on additional debt, and with the proceeds of this borrowing takes on more assets.

Figure 3.4 plots the change in assets against change in collateralized borrowing. The positive relationship in the scatter plot confirms our panel regression finding that balance sheet changes are accompanied by changes in short term borrowing.

Figure 3.5 plots the change in repos against the change in reverse repos. A dealer taking a short position by selling a security it does not own needs to deliver the security to the buyer on the settlement date. This can be done by borrowing

Figure 3.4:

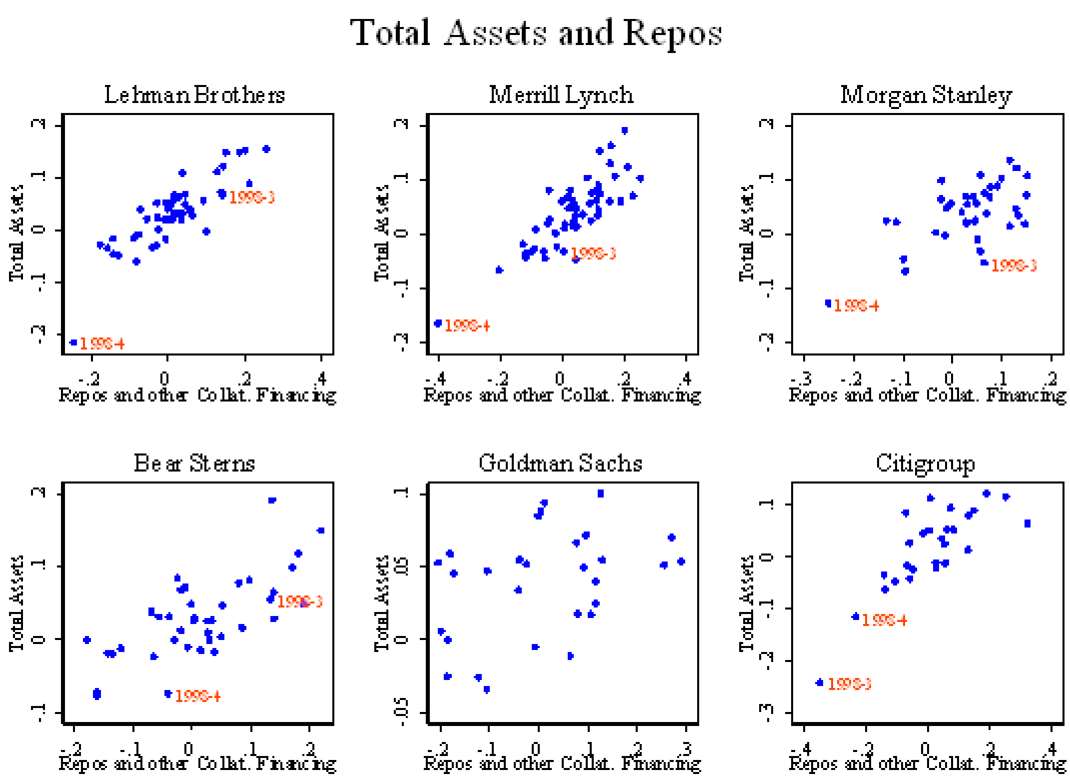
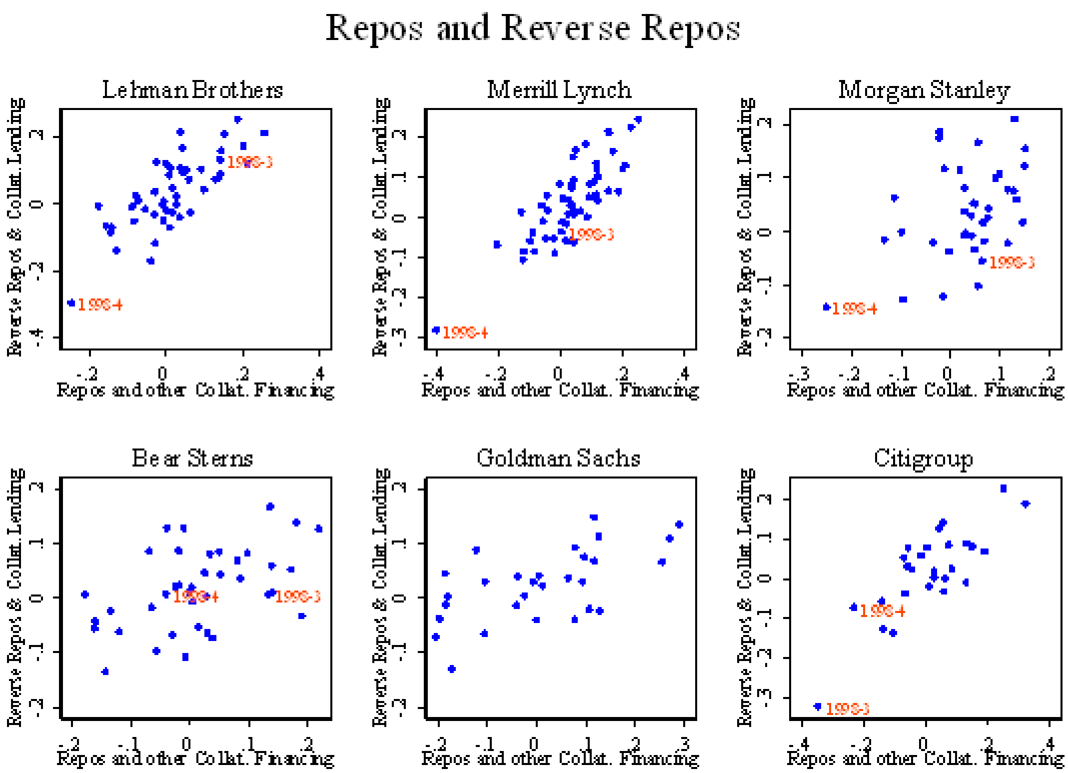

Total Assets and Repos
Lehman Brothers
Merrill Lynch
Morgan Stanley
Bear Sterns
Goldman Sachs
Citigroup
Total Assets
Repos and other Collat. Financing
1998-3
1998-4

Figure 3.5:


Repos and Reverse Repos
Lehman Brothers
Merrill Lynch
Morgan Stanley
Bear Sterns
Goldman Sachs
Citigroup
Reverse Repos & Collat. Lending
Repos and other Collat. Financing

the needed security, and providing cash or other securities as collateral. When the dealer closes out the short position by buying the security, the borrowed security can be returned to the securities lender. The scatter plot in figure 3.5 suggests that repos and reverse repos play such a role as counterparts in the balance sheet.

3.2. Value at Risk

Procyclical leverage is not a term that the banks themselves are likely to use in describing what they do, although this is in fact what they are doing. To get a better handle on what motivates the banks in their actions, we explore the role of value at risk (VaR) in explaining the banks' balance sheet decisions.

For a random variable A, the *value at risk* at confidence level c relative to some base level A_0 is defined as the smallest non-negative number VaR such that

$$\text{Prob}\left(A < A_0 - VaR\right) \leq 1 - c$$

For instance, A could be the total marked-to-market assets of the firm at some given time horizon. Then the value at risk is the equity capital that the firm must hold in order to stay solvent with probability c. Financial intermediaries publish their value at risk numbers as part of their regulatory filings, and also regularly disclose such numbers through their annual reports. Their *economic capital* is tied to the overall value at risk of the whole firm, where the confidence level is set at a level high enough to target a given credit rating (typically A or AA).

If financial intermediaries adjust their balance sheets to target a ratio of Value-at-Risk to economic capital, then we may conjecture that their disclosed Value-at-Risk figures would be informative in reconstructing their actions. If the bank maintains capital K to meet total value at risk, then we have

$$K = \lambda \times VaR \tag{3.1}$$

where λ is the proportion of capital that the intermediary holds per unit of VaR. The proportionality λ is potentially time varying. Hence, leverage L satisfies

$$L = \frac{A}{K} = \frac{1}{\lambda} \times \frac{A}{VaR}$$

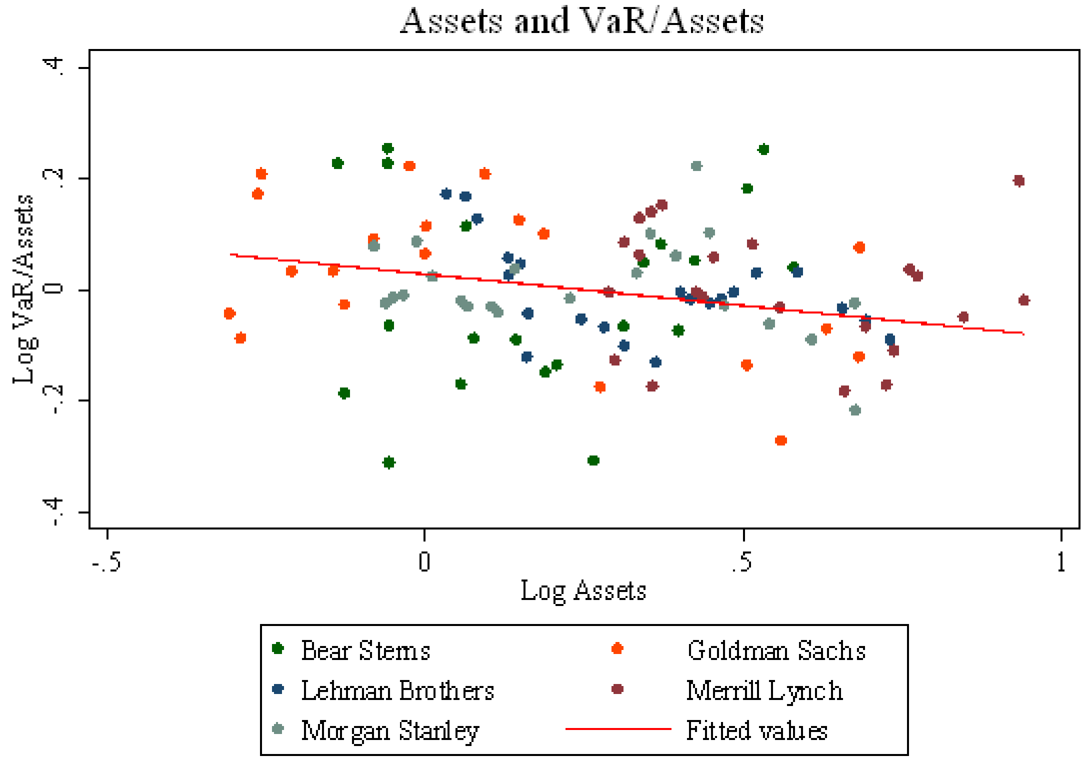
Procyclical leverage then translates directly to *counter*-cyclical nature of unit value-at-risk (i.e. value-at-risk per dollar of assets). Measured risk is low during booms and high during busts. We can indeed see this counter-cyclical relationship in the data. In Figure 3.6, we plot the VaR to total asset ratio against total assets and see that it is downwardsloping (we have removed fixed effects to produce this plot).

We explore the way in which the ratio of total value at risk to equity varies over time. Equation (3.1) suggests that it would be informative to track the ratio of value at risk to shareholder equity over time. The naive hypothesis would be that this ratio is kept constant over time by the bank. The naive hypothesis also ties in neatly the regulatory capital requirements under the 1996 Market Risk Amendment of the Basel capital accord. Under this rule, the regulatory capital is 3 times the 10 day, 99% value at risk. If total value risk is homogenous of degree 1, then (3.1) also describes the *required* capital for the bank, also.

In Figure 3.7 we plot the evolution of the VaR/equity ratio and leverage over time. We can see that both ratio are fairly constant. Only Goldman Sachs exhibits a marked increase in leverage (and a corresponding increase in VaR/Equity) over time. On average, both leverage and VaR/equity appear stationary, which is in accordance with the risk management and regulatory constraints.

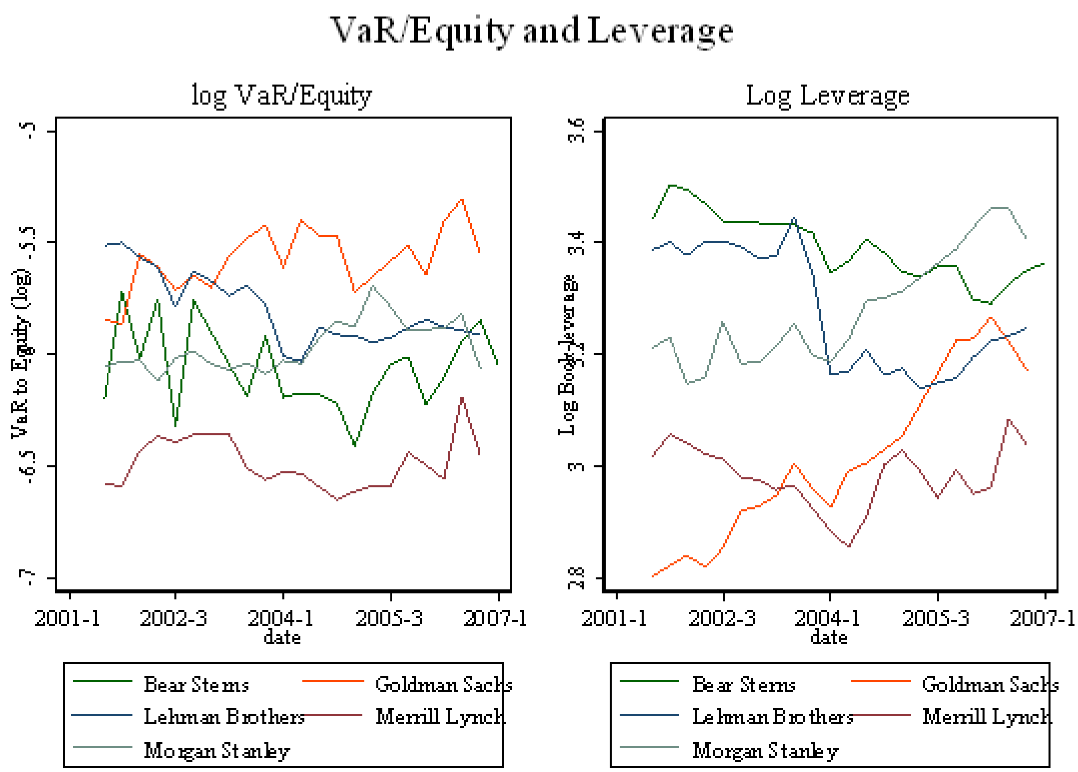
Table 4 presents the regressions for the quarterly change in the ratio of value at risk to equity. Value at risk numbers are those numbers that the banks themselves have reported in their 10-K and 10-Q filings. For the reasons outlined already, the firm's self-assessed value at risk is closely tied to its assessment of economic

Figure 3.6:



Assets and VaR/Assets
Log VaR/Assets
.4
.2
0
-.2
-.4
-.5
0
.5
1
Log Assets
Bear Sterns
Goldman Sachs
Lehman Brothers
Merrill Lynch
Morgan Stanley
Fitted values

Figure 3.7:



VaR/Equity and Leverage
log VaR/Equity
VaR to Equity (log)
-5
-5.5
-6.5
-7
2001-1
2002-3
2004-1
2005-3
2007-1
date
Log Leverage
Log Book-leverage
3.6
3.4
3
2.8
Bear Sterns
Goldman Sachs
Lehman Brothers
Merrill Lynch
Morgan Stanley

capital, and we would expect behavior to be heavily influenced by changes in value at risk.

[Table 4]

We focus on the ratio of value at risk to equity. In the panel regressions, the lagged value at risk to equity ratio is strongly negative, with coefficients in the range of -0.5 to -0.6, suggesting rapid reversion to the mean. We take this as evidence that the banks use VaR as a cue for how they adjust their balance sheets. However, the naive hypothesis that banks maintain a fixed ratio of value at risk to equity does not seem to be supported in the data. Column (ii) of Table 4 suggests that an increase in the value at risk to equity ratio coincides with periods when the bank increases its leverage. Value at risk to equity is procyclical, when measured relative to leverage. However, total assets have a negative sign in column (v). It appears that value at risk to equity is procyclical, but total assets adjust down some of the effects captured in leverage. The evidence points to an additional, procyclical risk appetite component to banks' exposures that goes beyond the simple hypothesis of targeting a normalized value at risk measure.

4. Forecasting Risk Appetite

We now present the main results of our paper. We show the asset pricing consequences of balance sheet expansion and contraction. We have already noted how the demand and supply responses to price changes can become perverse when financial intermediaries' actions result leverage that co-vary positively with the financial cycle. We exhibit empirical evidence that the waxing and waning of balance sheets have a direct impact on asset prices through the ease with which traders, hedge funds and other users of credit can obtain funding for trades.

So far, we have used quarterly data drawn either from the balance sheets of individual financial intermediaries or the aggregate balance sheet items from the Flow of Funds accounts. However, for the purpose of tracking the financial market consequences of balance sheet adjustments, data at a higher frequency is more likely to be useful. For this reason, we use the weekly data on the primary dealer repo and reverse repo positions compiled by the Federal Reserve Bank of New York.

Primary dealers are the dealers with whom the Federal Reserve has an on-going trading relationship in the course of daily business. The Federal Reserve collects data that cover transactions, positions, financing, and settlement activities in U.S. Treasury securities, agency debt securities, mortgage-backed securities (MBS), and corporate debt securities for the primary dealers. The data are used by the Fed to monitor dealer performance and market conditions, and are also consolidated and released publicly on the Federal Reserve Bank of New York website[5]. The dealers supply market information to the Fed as one of several responsibilities to maintain their primary dealer designation and hence their trading relationship with the Fed. It is worth noting that the dealers comprise an important but limited subset of the overall market. Moreover, dealer reporting entities may not reflect all positions of the larger organizations. Nevertheless, the primary dealer data provide a valuable window on the overall market, at a frequency (every week) that is much higher than the usual quarterly reporting cycle.

Dealers gather information at the close of business each Wednesday, on their transactions, positions, financing, and settlement activities over the previous week. They report on U.S. Treasury securities, agency debt securities, mortgage backed securities, and corporate debt securities. Data are then submitted on the following day (that is, Thursday) via the Federal Reserve System's Internet Electronic Sub-

[5] www.newyorkfed.org/markets/primarydealers.html

mission System. Summary data are released publicly by the Fed each Thursday, one week after they are collected. The data are aggregated across all dealers, and are only available by asset class (that is, Treasuries, agencies, etc.). Individual issue data, and individual dealer data, are not released publicly.

Repos and reverse repos are an important subset of the security financing data. The financing is reported on a gross basis, distinguishing between "securities in" and "securities out" for each asset class. "Securities in" refer to securities received by a dealer in a financing arrangement (be it against other securities or cash), whereas "securities out" refer to securities delivered by a dealer in a financing arrangement (be it against securities or cash). For example, if a dealer enters into a repo, in which it borrows funds and provides securities as collateral, it would report securities out. Repos and reverse repos are reported across all sectors. The actual financing numbers reported are the funds paid or received. In the case of a repo, for example, a dealer reports the actual funds received on the settlement of the starting leg of the repo, and not the value of the pledged securities. In cases where only securities are exchanged, the market value of the pledged securities is reported.

[Table 5]

We use the weekly repo and reverse repo data to forecast financial market conditions in the following week. Summary statistics are in Table 5. Our measure of financial market conditions is the VIX index of the weighted average of the implied volatility in the S&P500 index options. The VIX index has found widespread application in empirical work as a proxy for market risk appetite. Ang, Hodrick, Xing, and Zhang (2006) show that VIX innovations are significant pricing factors for the cross section of equity returns, and Bollerslev and Zhou (2007) show that the volatility risk premium —the difference between the VIX and realized volatility of the S&P500 index — forecasts equity returns better than other commonly used forecasting variables (such as the P/E ratio or the term spread).

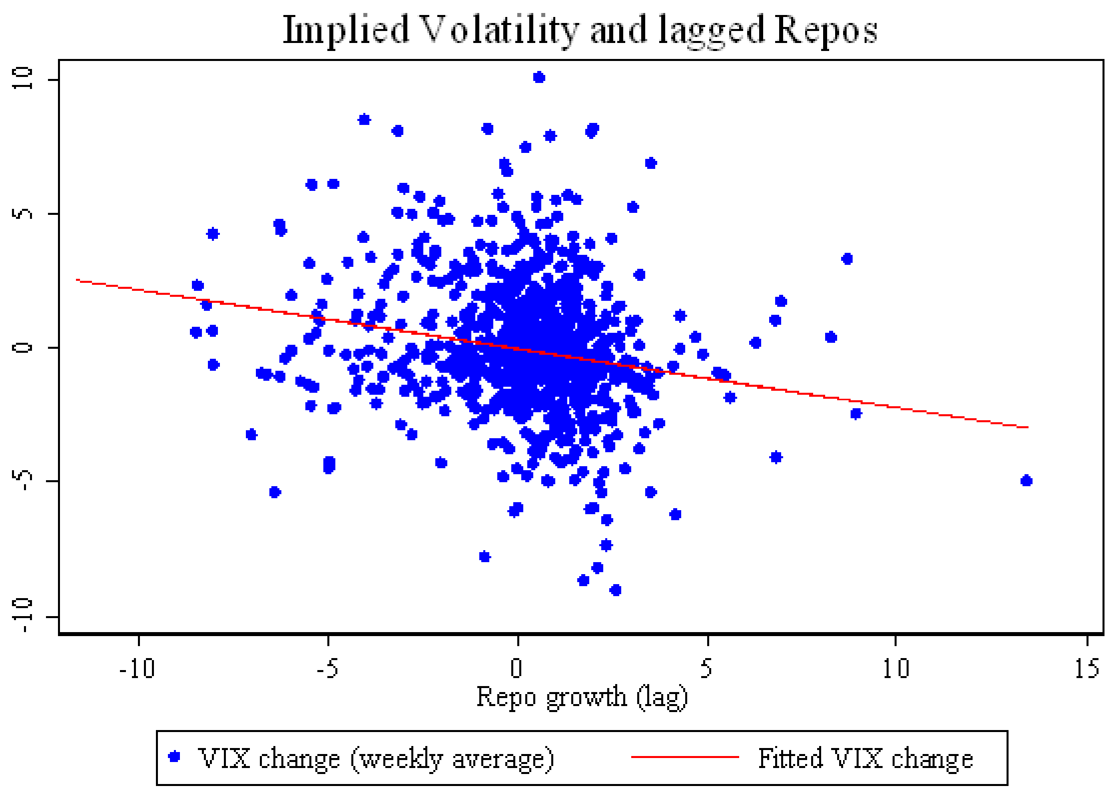
We use the daily VIX data from the website of the Chicago Board Options Exchange (www.cboe.com/micro/vix), and compute the S&P500 volatility from daily data over weekly windows. We compute the volatility risk premium as the difference between implied volatility and realized volatility. This risk premium is closely linked to the payoff to volatility swaps, which are zero investment derivatives that return the difference between realized future volatility and implied volatility over the maturity of the swap (see Carr and Wu (2007) for an analysis of variance and volatility swaps). We then compute averages of the VIX and the variance risk premium over each week (from the close of Wednesday to the close of the following Tuesday).

We are able to forecast innovations in the VIX. This can be seen in columns (ii)-(vi) of Table 6. We report forecasting regressions for VIX changes over the next week, as well as the Wednesday-Thursday and Wednesday-Friday changes. All of the forecasting results are significant at the 1% level. The forecasting R^2 increases from 8.9% when only the past VIX level is used, column (i) to 11.6% when Repo changes are included in the forecast. We believe the latter result (the ability to forecast the innovation in implied volatility) to be a very significant result. The forecasting result also holds for reverse repos, consistent with the notion that it is the total size of the balance sheet that matters for aggregate liquidity.

[Table 6]

In order to gain a better understanding what is determining the forecasting result, we also run the forecasting regressions for S&P500 volatility and the volatil-

Figure 4.1:



ity risk premium (columns vii-x). We see that it is the volatility risk premium that is being forecast, not actual equity volatility. Adjustments to the size of financial intermediary balance sheets via repos thus forecasts the price of risk of aggregate volatility, rather than aggregate volatility itself. We provide a graphical illustration of the forecasting power of repos in Figure 4.1.

We can put forward the following economic rationale for the forecasting regressions presented here. When balance sheets expand through the increased collateralized lending and borrowing by financial intermediaries, the newly released funding resources then chase available assets for purchase. More capital is deployed in increasing trading positions through the chasing of yield, and the selling of the "tails", as in the selling of out of the money puts. If the increased funding for asset purchases result in the generalized increase in prices and risk appetite in the financial system, then the expansion of balance sheets will eventually be reflected in the asset price changes in the financial system - hence, the ability of changes in repo positions to forecast future risk appetite.

5. Related Literature

The targeting of leverage seems closely to the bank's attempt to target a particular credit rating. To the extent that the "passive" credit rating should fluctuate with the financial cycle, the fact that a bank's credit rating remains constant through the cycle suggests that banks manage their leverage actively, so as to shed exposures during downturns. Kashyap and Stein (2003) draw implications from such behavior for the pro-cyclical impact of the Basel II bank capital requirements.

To the extent that balance sheets play a central role in our paper, our discussion here is related to the large literature on the amplification of financial shocks. The literature has distinguished two distinct channels. The first is the increased credit that operates through the *borrower's* balance sheet, where increased lending comes from the greater creditworthiness of the borrower (Bernanke and Gertler (1989), Kiyotaki and Moore (1998, 2001)). The second is the channel that operates through the *banks'* balance sheets, either through the liquidity structure of the banks' balance sheets (Bernanke and Blinder (1988), Kashyap and Stein (2000)), or the cushioning effect of the banks' capital (Van den Heuvel (2002)). Our discussion is closer to the latter group in that we also focus on the intermediaries' balance sheets. However, the added insight from our discussions is on the way that marking to market enhances the role of market prices, and the responses that price changes elicit from intermediaries.

Our results also related to the developing theoretical literature on the role of liquidity in asset pricing (Gromb and Vayanos (2002), Allen and Gale (2004), Acharya and Pedersen (2005), Brunnermeier and Pedersen (2005a, 2005b), Morris and Shin (2004), Acharya, Shin and Yorulmazer (2007a, 2007b)). The common thread is the relationship between funding conditions and the resulting market prices of assets. The theme of financial distress examined here is also closely related to the literature on liquidity drains that deal with events such as the stock market crash of 1987 and the LTCM crisis in the summer of 1998. Gennotte and Leland (1990) and Geanakoplos (2003) provide analyses that are based on competitive equilibrium.

The impact of remuneration schemes on the amplifications of the financial cycle have been addressed recently by Rajan (2005). The agency problems within a financial institution holds important clues on how we may explain procyclical behavior. Stein (1997) and Scharfstein and Stein (2000) present analyses of the capital budgeting problem within banks in the presence of agency problems.

The possibility that a market populated with value at risk (VaR) constrained traders may have more pronounced fluctuations has been examined by Danielsson, Shin and Zigrand (2004). Mark-to-market accounting may at first appear to be an esoteric question on measurement, but we have seen that it has potentially important implications for financial cycles. Plantin, Sapra and Shin (2005) present a microeconomic model that compares the performance of marking to market and historical cost accounting systems.

6. Concluding Remarks

Aggregate liquidity can be understood as the rate of growth of aggregate balance sheets. When financial intermediaries' balance sheets are generally strong, their leverage is too low. The financial intermediaries hold surplus capital, and they will attempt to find ways in which they can employ their surplus capital. In a loose analogy with manufacturing firms, we may see the financial system as having "surplus capacity". For such surplus capacity to be utilized, the intermediaries must expand their balance sheets. On the liabilities side, they take on more short-term debt. On the asset side, they search for potential borrowers that they can lend to. Aggregate liquidity is intimately tied to how hard the financial intermediaries search for borrowers. In the sub-prime mortgage market in the United States we have seen that when balance sheets are expanding fast enough, even borrowers that do not have the means to repay are granted credit - so intense is the urge to employ surplus capital. The seeds of the subsequent downturn in the credit cycle are thus sown.

References

Adrian, T. and H. S. Shin (2006) "Money, Liquidity and Financial Cycles" paper prepared for the Fourth ECB Central Banking Conference, "The Role of Money: Money and Monetary Policy in the Twenty-First Century", Frankfurt, November 9-10, 2006.

Allen, F. and D. Gale (2004) "Financial Intermediaries and Markets," *Econometrica* 72, 1023-1061.

Acharya, Viral and Lasse Pedersen (2005) "Asset Pricing with Liquidity Risk" *Journal of Financial Economics* 77, 375-410.

Ang, A., R. Hodrick, Y Xing, and X. Zhang (2006), "The Cross-Section of Volatility and Expected Returns," *Journal of Finance* 61, pp. 259-299.

Bank for International Settlements (1999): “A Review of Financial Market Events in Autumn 1998,” CGFS Publication Number 12, Bank for International Settlements, http://www.bis.org/publ/cgfs12.htm.

Bernanke, B. and A. Blinder (1988) “Credit, Money and Aggregate Demand” *American Economic Review*, 78, 435-39.

Bernanke, B. and M. Gertler (1989) “Agency Costs, Net Worth, and Business Fluctuations” *American Economic Review*, 79, 14 - 31.

Bollerslev, T. and H. Zhou (2007) ”Expected Stock Returns and Variance Risk Premia,” *Federal Reserve Board Finance and Discussion Series* 2007-11.

Brunnermeier, Markus and Lasse Heje Pedersen (2005a) “Predatory Trading”, *Journal of Finance*, 60, 1825-1863

Brunnermeier, Markus and Lasse Heje Pedersen (2005b) “Market Liquidity and Funding Liquidity”, working paper, Princeton University and NYU Stern School.

Jon Danielsson, Hyun Song Shin and Jean-Pierre Zigrand, (2004) “The Impact of Risk Regulation on Price Dynamics”, *Journal of Banking and Finance*, 28, 1069-1087

Diamond, Douglas and Raghuram Rajan (2005) “Liquidity Shortages and Banking Crises” *Journal of Finance*, 60, pp. 615

Geanakoplos, J. (2003) “Liquidity, Default, and Crashes: Endogenous Contracts in General Equilibrium” *Advances in Economics and Econometrics: Theory and Applications, Eighth World Conference, Volume II,* Cambridge University Press

Gennotte, Gerard and Hayne Leland (1990) “Hedging and Crashes”, *American Economic Review*, 999–1021.

Gromb, Denis and Dimitri Vayanos (2002) “Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs”, *Journal of Financial Economics*, 2002, 66, 361-407

Kashyap, A. and J. Stein (2000) “What Do a Million Observations on Banks Say about the Transmission of Monetary Policy?” *American Economic Review*, 90, 407-428.

Anil Kashyap and Jeremy Stein, 2003, “Cyclical Implications of the Basel II Capital Standard”, University of Chicago, Graduate School of Business and Harvard University, http://faculty.chicagogsb.edu/anil.kashyap/research/basel-final.pdf

Kiyotaki, N. and J. Moore (1998) “Credit Chains” LSE working paper, http://econ.lse.ac.uk/staff/kiyotaki/creditchains.pdf.

Kiyotaki, N. and J. Moore (2001) “Liquidity and Asset Prices” LSE working paper, http://econ.lse.ac.uk/staff/kiyotaki/ liquidityandassetprices.pdf.

Carr, P. and Wu, L., (2007) ”Variance Risk Premia,” *Review of Financial Studies*, forthcoming.

Lowenstein, R. (2000) *When Genius Failed*, Random House, New York.

Plantin, G., H. Sapra and H. S. Shin (2005) “Marking to Market: Panacea or Pandora’s Box” working paper, Princeton University

Rajan, R. (2005) “Has Financial Development Made the World Riskier?” paper presented at the Federal Reserve Bank of Kansas City Economic Symposium at Jackson Hole,
http://www.kc.frb.org/publicat/sympos/2005/sym05prg.htm

Scharfstein, David and Jeremy Stein (2000) "The Dark Side of Internal Capital Markets: Divisional Rent-Seeking and Inefficient Investment" *Journal of Finance*, 55, 2537-2564.

Stein, Jeremy (1997) "Internal Capital Markets and the Competition for Corporate Resources" *Journal of Finance*, 52, 111-133.

Van den Heuvel, S. (2002) "The Bank Capital Channel of Monetary Policy," working paper, Wharton School, University of Pennsylvania,
http://finance.wharton.upenn.edu/˜vdheuvel/BCC.pdf

Table 2: Investment Bank Summary Statistics

This Table reports aggregate balance sheet items for the five investment banks of Table 1. In Panel A, we report time series summary statistics for the cross sectional average of the balance sheet items. In Panel B, we report the summary statistics of quarterly grwoth rates which are weighted by the Total Assets cross sectionally.

Panel A: US$ Millions	Mean	Std Dev	Min	Median	Max	Obs
Total Assets	301460	163696	97302	265079	730825	60
Total Liabilities	288739	157018	93111	254984	702510	60
Equity	11908	7172	3426	9246	28302	60
Reverse Repos and other Collateralized Lending	94222	46691	29423	86515	217254	60
Reverse Repos	58612	24191	19097	54028	125601	60
Repos and other Collateralized Borrowing	120139	64681	34216	114162	282272	60
Repos	88899	31491	54682	80030	169110	48
Trading VaR	49	13	29	47	82	24

Panel B: Quarterly Growth	Mean	Std Dev	Min	Median	Max	Obs
Total Assets	4%	5%	-15%	4%	16%	59
Total Liabilities	4%	6%	-15%	4%	17%	59
Equity	3%	2%	-2%	4%	10%	59
Reverse Repos and other Collateralized Lending	3%	9%	-26%	4%	21%	59
Reverse Repos	3%	9%	-16%	2%	28%	59
Repos and other Collateralized Borrowing	4%	7%	-19%	3%	21%	59
Repos	2%	9%	-19%	1%	19%	48
Trading VaR	3%	8%	-17%	3%	19%	23

Table 3: Explaining Leverage

This table reports panel regressions of quarterly leverage growth rates on the lagged level of leverage, the growth rates of trading VaRs, the growth rates of repos, and the growth rates of total assets. Leverage is computed from the balance sheets of the five investment banks from Table 1 whose summary statistics are reported in Table 2. Leverage is defined as the ratio of total assets to book equity. All of the balance sheet data is from the 10-K and 10-Q filings of the banks with the Security and Exchange Commission, and is taken from the Mergent Database.

		Leverage (quarterly growth)			
		(i)	(ii)	(iv)	(v)
Leverage (log lag)	coef	-0.09	-0.10	-0.04	-0.04
	p-value	0.00	0.01	0.03	0.00
Trading VaR (quarterly growth)	coef		0.07		
	p-value		0.02		
Repos (quarterly growth)	coef			0.37	
	p-value			0.00	
Total Assets (quarterly growth)	coef				0.90
	p-value				0.00
Constant	coef	0.28	0.32	0.12	0.10
	p-value	0.00	0.01	0.04	0.01
Observations		211	108	211	211
Number of Banks		5	5	5	5
R-squared		5%	12%	43%	66%
Fixed Effects		yes	yes	yes	yes

Table 4: Explaining the VaR/Equity Ratio

This table reports panel regressions of quarterly growth rates of the ratio of VaR to equity on the lagged level of leverage, the growth rates of trading VaRs, and the growth rates of total assets. The data is for the five investment banks from Table 1 whose summary statistics are reported in Table 2. All of the balance sheet data is from the 10-K and 10-Q filings of the banks with the Security and Exchange Commission, and is taken from the Mergent Database.

		Trading VaR / Equity (quarterly growth)			
		(i)	(ii)	(iii)	(iv)
Trading VaR / Equity (log lag)	coef	-0.61	-0.56	-0.62	-0.54
	p-value	0.00	0.00	0.00	0.00
Leverage (quarterly growth)	coef		0.91		1.65
	p-value		0.00		0.00
Total Assets (quarterly growth)	coef			-0.04	-1.29
	p-value			0.90	0.00
Constant	coef	-3.67	-3.32	-3.68	-3.20
	p-value	0.00	0.00	0.00	0.00
Observations		107	107	107	107
Number of i		5	5	5	5
R-squared		33%	39%	33%	44%
Fixed Effects		yes	yes	yes	yes

Table 5: Primary Dealer Financing Summary Statistics

This Table reports summary statistics of collateralized financing by the Federal Reserve's Primary Dealers from form FR2004 for January 3, 1990 - August 29, 2007.

Panel A: US$ Billions	Mean	Std Dev	Min	Max	Obs
Reverse Repos and other Collateralized Lending	1712	1010	382	4076	896
Reverse Repos	1655	1008	369	4040	896
Repos and other Collateralized Borrowing	1636	961	397	3896	896
Repos	1204	663	332	2636	896
Net Repos	451	357	21	1456	896

Panel B: Weekly Growth	Mean	Std Dev	Min	Max	Obs
Reverse Repos and other Collateralized Lending	18%	217%	-1092%	1360%	895
Reverse Repos	19%	223%	-1162%	1344%	895
Repos and other Collateralized Borrowing	17%	209%	-1097%	1266%	895
Repos	19%	264%	-1388%	1471%	895
Net Repos	40%	443%	-2429%	5356%	895

Table 6: Forecasting Volatility

This table reports forecasting regressions of VIX implied volatility changes, S&P500 volatility changes, and the volatility risk premium on lagged growth rates of repo, reverse repo, and net repo positions of U.S. Primary Dealers. The VIX is computed from the cross section of S&P500 index option prices by the Chicago Board of Options Exchange. We compute weekly volatility from S&P500 returns. the volatility risk premium is the difference between the average VIX over the week and S&P500 volatility for the same week. Summary statistics of the Primary Dealer financing data are given in Table 5. The data is weekly from January 3, 1990 - August 29, 2007. P-values are adjusted for autocorrelation and heteroskedasticity.

		Implied Volatility (Change)							Volatility (Change)		Volatility Risk Premium	
		One week average				Wed-Thur	Wed-Fri	Thur-Fri				
		(i)	(ii)	(iii)	(iv)	(v)	(vi)	(vi)	(vii)	(viii)	(ix)	(x)
Implied Volatility	coef	-0.12	-0.11	-0.11	-0.12	-0.01	-0.03	-0.03	-0.45	-0.45	0.22	0.21
(lag)	p-value	0.00	0.00	0.00	0.00	0.15	0.00	0.00	0.00	0.00	0.00	0.00
Repos	coef		-0.20			-0.05	-0.05	-0.05		0.05		-0.16
(lagged growth)	p-value		0.00			0.01	0.04	0.04		0.52		0.03
Reverse Repos	coef			-0.14								
(lagged growth)	p-value			0.00								
Net Repos	coef				-0.06							
(lagged growth)	p-value				0.00							
Constant	coef	2.16	2.09	2.09	2.14	0.16	0.38	0.38	4.93	4.90	6.23	6.30
	p-value	0.00	0.00	0.00	0.00	0.19	0.03	0.03	0.00	0.00	0.00	0.00
Observations		903	878	878	878	878	878	878	878	878	878	878
R-squared		8.9%	11.6%	10.9%	10.1%	1.1%	1.6%	1.6%	22.8%	22.0%	40.2%	40.9%

Liquidity and Congestion*

Gara Mínguez Afonso

London School of Economics
gara@princeton.edu

May 14, 2008

Abstract

This paper studies the relationship between the arrival of potential investors and market liquidity in a search-based model of asset trading. The entry of investors into a specific market causes two contradictory effects. First, it reduces trading costs, which then attracts new investors (thick market externality effect). But secondly, as investors concentrate on one side of the market, the market becomes "congested", decreasing the returns to participating in this market and discouraging new investors from entering (*congestion effect*). The equilibrium level of market liquidity depends on which of the two effects dominates. When congestion is the leading effect, some interesting results arise. In particular, we find that diminishing trading costs in our market can deteriorate liquidity and reduce welfare.

*I am indebted to Hyun Song Shin for continuing guidance and encouragement. For very valuable advice, I sincerely thank Dimitri Vayanos and Markus Brunnermeier. I am grateful to Jean-Charles Rochet for a constructive discussion of the paper, to Pierre-Olivier Weill for helpful suggestions and to Max Bruche, Julio Cacho-Díaz, Nicola Cetorelli, Camille Cornand, Amil Dasgupta, Nobuhiro Kiyotaki, Péter Kondor, Albert "Pete" Kyle, Paolo Pasquariello, Ashley Taylor, Laura Veldkamp, S. "Vish" Viswanathan, Wei Xiong and participants at the Cycles, Contagion and Crises Conference, the Princeton Finance Research Camp and seminars at Princeton University, University of Lausanne, McIntire School of Commerce (University of Virginia), Fuqua School of Business (Duke University), Stephen M. Ross School of Business (University of Michigan), Federal Reserve Bank of New York, Robert H. Smith School of Business (University of Maryland), the Board of Governors of the Federal Reserve System and the Workshop in Economic Theory at UCLA for helpful comments. Also, I am thankful to the Bendheim Center for Finance (Princeton University) for their hospitality, to the Economic and Social Research Council for its support under research grant RES-156-25-0026 and to Fundación Ramón Areces for financial support.

1 Introduction

Liquidity is sometimes defined as a coordination phenomenon. In financial markets, as investors move into a specific market they facilitate trade for all investors by reducing the cost of participating in this market. At the same time, easier trade and lower trading costs attract potential investors. There is a thick market externality where new investors provide market liquidity and market liquidity attracts new investors. However, as investor prefer to join one side of a market, i.e. as they become buyers or sellers, this side of the market becomes "congested", hindering trade. Congestion then discourages investors from entering this market.

One-sided markets arise during financial booms and, more drastically, during market crashes. When a market is in distress, liquidity typically vanishes playing a key role in the build-up of one-sided markets. The study of liquidity in one-sided markets is thus vital to understand the response of financial systems to the threat of market disruptions. Recent episodes of market distress include the LTCM crisis[1] in 1998, the September 11, 2001, events[2] and the turbulence in credit markets[3] during the summer of 2007.

In this paper we present an alternative view of market liquidity. The main difference with the previous literature is that we consider not only a thick market externality but also a *congestion effect*. In our model, the arrival of new investors causes two opposite effects. First, it diminishes transaction costs and eases trade, which attracts potential investors. But secondly, as investors concentrate on one side of the market, trade becomes more difficult, reducing the returns to participating in this market and discouraging potential investors from entering. Market liquidity thus results from the tradeoff between thick market externalities and a congestion effect.

We assume an infinite-horizon steady-state market where agents can invest in one asset which can be traded only bilaterally. In this market, investors cannot trade instantaneously

[1]For an analysis of the events surrounding the market turbulence in autumn 1998, see BIS (1999) and IMF (1998).

[2]Cohen and Remolona (2001) presents a summary of the September 11, 2001 episode in global financial markets. Also, McAndrews and Potter (2002) gives a detailed account of the consequences of the September 11, 2001, events on the US payment system and of the actions of the Federal Reserve System to provide liquidity to the financial system.

[3]See Fender and Hördahl (2007) for an overview of the key events over the period from end-May to end-August 2007.

but it takes some time to find a trading partner resulting in opportunity and other costs. Once an investor buys the asset, he holds it until his preference for the ownership change and he prefers to liquidate the investment and exit the market. To model the search process we adopt the framework introduced in Vayanos and Wang (2007). In our setting though, investors are heterogeneous in their investment opportunities in the sense that some investors have access to better investment options than others.

We compute explicitly the unique equilibrium allocations and the price at which investors trade with each other and show how they depend on the flow of new investors entering the market. Prices negotiated between investors are higher in the flow of potential investors. However, investors' entry decision is endogenous and thus depends on market, asset and investors characteristics. A change in investors' search abilities, for instance, affects both the rate of meetings between trading partners and the flow of investors entering the market, which then determines the distribution of potential partners with whom they can meet.

Moreover, the equilibrium flow of investors arises from a tradeoff between thick market complementarities and a congestion effect. When congestion is the dominating effect some interesting results come to light. First, reducing market frictions can *decrease* market attractiveness. Under some cases, one-sided markets can develop. A regulatory reform or the introduction of a technological advance, such as a new electronic trading system, can induce an adverse effect on the distribution of investors during upswings. Specifically, it would allow the few sellers present in the market to exit at faster rates leading to an even more unbalanced distribution of investors. Congestion then intensifies as the market becomes more one-sided, discouraging potential investors and thus dampening down the attractiveness of this market.

Second, *diminishing market frictions can deteriorate market liquidity and reduce welfare.* The reason for this counterintuitive result is the following. In a one-sided market with more sellers than buyers, for example during a fire sale, introducing a measure that improves the efficiency of the search process makes it easier for one of the few buyers present in the market to acquire the asset. But when the buyer purchases the asset (and a seller exits), the proportion of buyers to sellers falls further and the market becomes more one-sided. As investors cluster on the sell-side of this market, buyers gain a more favourable position in the bargaining process and try to lower the price they pay to acquire the asset. Reducing market frictions in a distressed market thus magnifies the effect of congestion and results in a lower asset price (a higher price discount) and ultimately in a less liquid market. Investors who hold this asset and those trying to sell it are clearly worse-off as the market becomes more one-sided, leading to a decrease in overall welfare. From this point of view, this paper provides an example of the Theory of the Second Best. Improving the efficiency of the search process, when there are other imperfections in the market such as the ones arising from the congestion effect, is not necessarily welfare enhancing.

Third, market illiquidity measured by the price discount can increase while trading volume rises. Reducing search frictions during downswings amplifies the effect of congestion, resulting in a higher price discount and in a less liquid market. But a more efficient search process also increases the frequency of meetings between the investors already present in the market. More frequent meetings then translates into a higher trading volume. Thereby, a measure intended to shorten the waiting times needed to locate a trading partner in a market experiencing distressed selling can cause both higher price discount and higher trading volume. This third result joins the discussion on the measurement of the effect of liquidity on asset prices and shows how alternative measures capture different dimensions of market liquidity.

The outline of the paper is as follows. In the next section, we discuss the related literature. We introduce a theoretical framework to examine the relationship between market liquidity and the arrival of potential investors to this market in Section 3. Section 4 determines the population of investors, their expected utilities and the price of the asset, taking as given investors' decision to enter the market. Then, Section 5 endogenises the entering rule and characterises the study of the unique market equilibrium. Market liquidity and welfare are discussed in Section 6. Finally, Section 7 concludes. Some proofs and additional results are presented in the appendices.

2 Related Literature

The notion of thick market complementarity is clearly captured in Diamond (1982a). He considers an economy where islanders face production opportunities and decide whether to remain unemployed or to climb a palm tree and retrieve coconuts. Trees differ in their heights (the cost of production). Islanders only climb trees shorter than a certain height and they cannot consume the coconuts they pick. They need to search for a trade to swap the coconuts. The likelihood of meeting a trading partner in this economy increases in the number of potential traders available. This key feature constitutes the basis of the strategic complementarity in Diamond's model. This is highlighted in Cooper and John (1988), where they discuss the economic relevance of strategic complementarities in agents' payoffs and explain how they can lead to coordination failures. A related argument is presented in Milgrom and Roberts (1990). They show the Diamond-type search model is a supermodular[4] game, where more production or participation activity by some islanders raises the returns to increased levels of activity by others.

Building on strategic complementarities Brunnermeier and Pedersen (2007) and Gromb and Vayanos (2002) analyse the link between capital and market liquidity. Also, Pagano (1989) focuses on the feedback loop between trading volume and liquidity to study concentration and fragmentation of trade across markets. In Dow (2004), multiple equilibria with different degrees of market liquidity result from informational asymmetries. Plantin (2004) assumes investors can learn privately about an issuer's credit quality by holding an asset. This "learning by trading" also creates a thick market externality. From a broad perspective, this literature studies liquidity as a self-fulfilling phenomenon where both liquid and illiquid market equilibria may arise. Illiquid markets are thus a consequence of a coordination failure.

Our paper is also related to the search literature. The economics of search have their roots in Phelps (1972). Search-theoretic models such as the frameworks introduced in labour markets[5] by Diamond (1982a), Diamond (1982b), Mortensen (1982) and Pissarides (1985) have been broadly used in different areas of economics. In asset pricing[6], Duffie, Gârleanu and Pedersen introduce search and bargaining in models of asset market equilibrium to study the impact of these sources of illiquidity on asset prices. This paper is related to Duffie et al. (2005), which presents a theory of asset pricing and marketmaking in over-the-counter markets with search-based inefficiencies. They conclude that risk neutral investors receive narrower bid-ask spreads if they have easier access to other investors and marketmakers. Similarly to Duffie et al. (2005) we consider risk-neutral agents who can only invest in one asset. In our model though, investors can only trade with other investors and our focus,

[4]In the unidimensional case, a supermodular game is a game exhibiting strategic complementarities in which each agent's strategy set is partially ordered. See Topkis (1979) and Cooper (1999) for a formal definition.

[5]See Pissarides (2001) for a review of the literature on search in labour markets.

[6]For an excellent review on liquidity and asset prices, see Amihud et al. (2005).

rather than on liquidity and marketmaking, lies on the endogenous relationship between market liquidity and the arrival of potential investors to this market.

Duffie et al. (2007) extends their setting to incorporate risk aversion and risk limits and finds that, under certain conditions, search frictions as well as risk aversion, volatility and hedging demand increase the illiquidity discount. Lagos and Rocheteau (2007) also generalises Duffie et al. (2005) to allow for general preferences, unrestricted long positions, idiosyncratic and aggregate uncertainty and entry of dealers. Our paper shares with theirs the existence of strategic complementarities and an endogenous entry decision. To define the entry of dealers, Lagos and Rocheteau (2007) specify that the contact rate between investors and dealers increases sublinearly in the number of dealers. In our framework, entry is the result of a decision problem where investors compare the benefits of this market to their best investment opportunities.

Weill (2007) and Vayanos and Wang (2007) extend the framework of Duffie, Gârleanu and Pedersen to allow investors to trade multiple assets[7]. They show that search frictions lead to cross-sectional variation in asset returns due to illiquidity differences. In Vayanos and Wang (2007) investors are heterogeneous in their trading horizons while in Weill (2007) investors are homogeneous, but there are differences in the assets' number of tradable shares. From a methodological point of view, our paper is closely related to Vayanos and Wang (2007). The main difference with their work is that we consider only one asset and focus on the analysis of the liquidity in the market for this asset rather than on the liquidity across two assets.

Our paper is close in spirit to Huang and Wang (2007). They also find that decreasing market frictions can diminish the level of market liquidity. However, their framework and the general mechanism that yields this result clearly differ from ours. Rather than a search-based model, they consider a centralised market where exogenous transaction costs take the form of participation costs. Agents can pay an ex-ante cost to trade constantly (and become market makers) or pay a spot cost to trade after observing their trading needs. Huang and Wang (2007) argues that, when there is insufficient supply of liquidity, lowering the cost to enter on the spot can decrease welfare because it reduces investors' incentives to become market makers. In our model, market liquidity results from a tradeoff between thick market

[7]See also Vayanos and Weill (2007) for an application to the on-the-run phenomenon, by which recently issued bonds have higher prices than older ones with the same cash flows. They develop a multi-asset model where both the spot market and the repo market operate through search.

externalities and congestion effects. We show that, when the congestion effect dominates, the market becomes one-sided and improving the efficiency of the search process can diminish market liquidity because it discourages agents from investing into our market.

This paper also relates to the literature on asset pricing with exogenous trading costs studied in Amihud and Mendelson (1986), Vayanos (1998) and Acharya and Pedersen (2005), among others. We complement this literature by endogenising transaction costs.

3 The Model

Time is continuous and goes from zero to infinity. There is only one asset traded in the market with a total supply S. This asset pays a dividend flow d.

Consider risk-neutral agents, whom we will refer to as investors. By assuming risk neutrality, we aim to study the concentration of liquidity in a specific market without reference to investors' shifts in their attitudes towards risk. Investors are infinitely lived and have time preferences determined by a constant discount rate equal to $r > 0$. At some random time, investors decide to enter the market and aim to buy one unit of the asset. They become *buyers-to-be*. Once they purchase the asset, buyers-to-be become *non-searcher investors*. Non-searchers hold the asset and enjoy the full value d of its dividend flow until they receive a liquidity shock which makes them want to liquidate their portfolio and leave the market. We assume liquidity shocks arrive with a Poisson rate γ and reduce investors' valuation to a lower level $d - x$ of flow utility, where $x > 0$ captures the notion of a liquidity shock to the investors, for example, a sudden need for cash or the arrival of a good investment opportunity in another market. x could also be understood as the holding cost borne by the investor who receives a liquidity shock and is aiming to exit the market. At that time, non-searcher investors become *sellers-to-be* and seek to sell[8]. Upon selling, investors exit the market and join the initial group of outside investors.

The flow of investors entering the economy is defined by a function f. Investors are heterogeneous in their investment opportunities κ, i.e. we consider they differ on their outside

[8]Investors are risk neutral and thus have linear utility over the dividend flow d. Consequently, they optimally prefer to hold a maximum long position in the asset (which we can normalise to 1) or zero units of the asset (once they seek to exit the market).

options as some investors enjoy better investment possibilities than others. We assume f is a continuous and strictly positive function of the investor's investment opportunity class κ, such that the total flow of investors entering the economy is given by $\int_{\underline{\kappa}}^{\overline{\kappa}} f(\kappa)d\kappa$, where $[\underline{\kappa}, \overline{\kappa}]$ is the support of $f(\kappa)$. Only a fraction $\nu(\kappa)$ of the flow of investors entering the economy decides to invest in this market. At any point in time there is a non-negative flow of every class of investor from the outside investors' group into the market, and hence the total flow of investors entering the market is defined by $g = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu(\kappa)f(\kappa)d\kappa$.

We assume markets operate through search, with buyers and sellers matched randomly over time in pairs. Search is characteristic of over-the-counter markets where investors need to locate trading partners and then bargain over prices. There is a cost associated to this search process. In a market where it is more likely to find a counterpart in a short time, the search cost is smaller and liquidity, measured by search costs, is higher. But we could think of a broader interpretation of the search friction. In a centralised market, it represents the cost of being forced to trade with an outside investor who does not understand the full value of the asset and requires an additional compensation for trading. These investors only buy the asset at a discount and sell it at a premium. This transaction cost decreases in the abundance of investors. In the market of a frequently traded asset, it is less likely that it is necessary to trade with an outside investor who "mis-values" the asset and hence the transaction cost linked to this asset is smaller and its liquidity higher. In this paper, we use the first intuition because of its more transparent interpretation.

We adopt the search framework presented in Vayanos and Wang (2007). To define the search process, we first need to describe the rate at which investors willing to buy meet those willing to sell and once they meet we need to specify how the asset price is determined. The ease in finding a trading partner depends on the availability of potential partners. Let us consider that an investor seeking to buy or sell meets other investors according to a Poisson process with a fixed intensity. Thus, for each investor the arrival of a trading partner occurs at a Poisson rate proportional to the measure of the partner's group. Denote by η_b the measure of buyers-to-be and by η_s the measure of investors seeking to sell (sellers-to-be). Then, a buyer-to-be meets sellers-to-be with a Poisson intensity $\lambda\eta_s$ and a seller-to-be meets buyers-to-be at a rate $\lambda\eta_b$, where λ measures the efficiency of the search and a high λ represents an efficient search process. The overall flow of meetings[9] between trading partners is then given by $\lambda\eta_b\eta_s$.

Once investors meet they bargain over the price p of the asset. These meetings always result in trade as Proposition 5 shows. For simplicity we assume that either the investor willing to buy or the one willing to sell is chosen randomly to make a take-it-or-leave-it offer to his trading partner. Denoting by $\frac{z}{1+z}$ the probability of the buyer-to-be being selected to make the offer and thus by $\frac{1}{1+z}$ the probability that the seller-to-be makes the offer, $z \in (0, \infty)$ captures the buyer's-to-be bargaining power.

Figure 1 describes our market, specifying the different types of investors and the flows between types. η_0 denotes the measure of non-searcher investors.

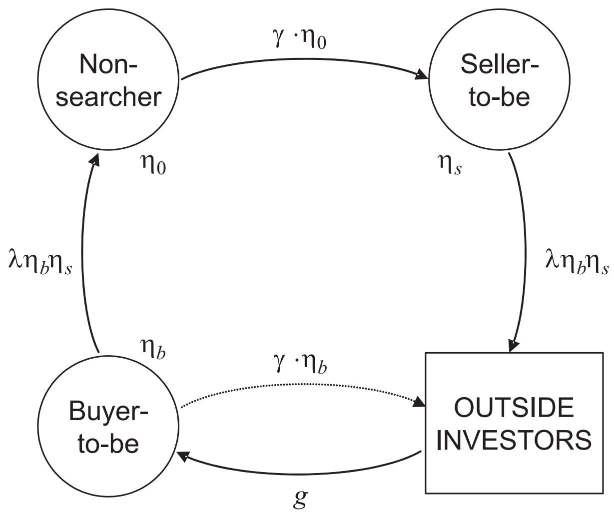


Figure 1: An outside investor enters the market and becomes a buyer-to-be aiming to meet a seller-to-be. If he suffers a liquidity shock before meeting a trading partner, he exits the market. On the contrary, if he meets a seller-to-be, he bargains over the price, buys the asset (pays p) and becomes a non-searcher. He holds the asset until he receives a liquidity shock. At that time, he becomes a seller-to-be seeking a buyer-to-be. When he meets a buyer-to-be, he bargains over the price, sells the asset (receives p) and exits the market returning to the group of outside investors.

[9]See Duffie and Sun (2007) for a formal proof of this result. This application of the exact law of large numbers for random search and matching has previously been used in Duffie et al. (2005), Duffie et al. (2007) and Vayanos and Wang (2007) among others.

4 Analysis

In this section we first solve for the steady-state measure of every type of investor in the market. Next, in Subsection 4.2, we describe investors' flow utilities. We show in Subsection 4.3 that every meeting between trading partners results in trade and we discuss thick market externalities in Subsection 4.4.

4.1 Measure of Investors

In this subsection we determine the measure of buyers-to-be (η_b), non-searcher investors (η_0) and sellers-to-be (η_s). Although investors are heterogeneous in their investment opportunities κ, once they enter the market their class does not alter their behaviour in this market. Investors develop sudden needs for cash at the same Poisson rate γ, independently of their outside investment opportunities κ. In consequence, we do not need to consider the distribution of investment opportunities within each population but the aggregate measure of buyers-to-be, non-searcher investors and sellers-to-be. This assumption could be generalised by considering γ a function of the outside option κ. The analysis would be similar but the notation more complicated, as we would need to take into account the distribution of investment opportunities κ within each group of investors rather than the aggregate measures[10].

In equilibrium, the market needs to clear and thus the supply of the asset equals the measure of investors holding the asset, each of whom holds one unit of the asset. Specifically, the sum of the measures of non-searchers and sellers-to-be is equal to the total supply of the asset:

$$\eta_0 + \eta_s = S \quad \Rightarrow \quad \eta_s = S - \eta_0 \tag{1}$$

In a steady state, the inflow of investors joining a group matches the outflow such that the rate of change of the group's population is zero. The inflow and outflow of the different types of investors are summarised in Figure 1. Let us first consider the non-searcher investors. In this case, inflows are given by the buyers-to-be who meet a trading partner and buy the asset ($\lambda\eta_b\eta_s$), while non-searchers receiving a liquidity shock ($\gamma\eta_0$) constitute the outflow. Setting

[10] See Section 3 in Vayanos and Wang (2007) for a particular case.

inflow equal to outflow and using equation (1) yields:

$$\eta_b = \frac{\gamma}{\lambda} \frac{\eta_0}{S - \eta_0} \tag{2}$$

We now analyse the population of buyers-to-be. The flows of investors coming from the outside group are defined by g. The outflow is comprised of the buyers-to-be who receive a liquidity shock before meeting a trading partner ($\gamma\eta_b$) and of those who meet sellers-to-be and buy the asset ($\lambda\eta_b\eta_s$). Then,

$$g = \gamma\eta_b + \lambda\eta_b\eta_s$$

Using equations (1) and (2) we can rewrite the previous equation as:

$$g = \gamma \left(1 + \frac{\gamma}{\lambda} \frac{1}{S - \eta_0}\right) \eta_0 \tag{3}$$

Equation (3) determines η_0 as a function of g. Then, substituting η_0 in equations (1) and (2) specifies η_s and η_b respectively. Let us first assume the flow of investors entering the market g is constant. We generalise our results in Subsection 5.1.

Proposition 1. *Given g constant, there is a unique solution to the system (1) - (3) given by:*

$$\eta_0 = \frac{1}{2\gamma} A \tag{4}$$

$$\eta_s = S - \frac{1}{2\gamma} A \tag{5}$$

$$\eta_b = \frac{\gamma}{\lambda} \frac{A}{2\gamma S - A} \tag{6}$$

where $A = (g + \gamma S + \frac{\gamma^2}{\lambda}) - \sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}$.

The proof is presented in Appendix A. It is interesting to note how the different measures of investors respond to changes in the parameters of our model. For instance, as the flow of investors g entering the market rises, the measure of investors willing to buy (buyers-to-be) and of those passively holding the asset (non-searchers) increase. However, given that there are more investors seeking to buy the asset, it is now easier for a seller-to-be to find a trading partner and hence the measure of investors seeking to sell falls. This is summarised in Proposition 2 and proven in Appendix A.

Proposition 2. *The measure of buyers-to-be and non-searcher investors is increasing in g* ($\frac{\partial \eta_b}{\partial g}, \frac{\partial \eta_0}{\partial g} > 0$) *while the measure of sellers-to-be decreases in g* ($\frac{\partial \eta_s}{\partial g} < 0$)*.*

Given the measures of investors η_b seeking to buy and those η_s seeking to sell, the efficiency of the search process λ defines the overall flow of meetings (and transactions, according to Proposition 5) in our market. However, the measures of the different types of investors also depend on the efficiency of the search process. In particular, for the same level of investors entering the market, if the search process is more efficient, there will be a lower measure of investors "waiting" to meet a potential seller ($\frac{\partial \eta_b}{\partial \lambda} < 0$). Thus, outside investors, who enter the market, meet a trading partner and become non-searcher investors at a faster rate if the search process is more efficient ($\frac{\partial \eta_0}{\partial \lambda} > 0$). A proportion of non-searcher investors then joins the pool of sellers-to-be and hence there is a higher flow of investors coming from the non-searchers to the group of sellers-to-be. And, although there are more inflows of investors and less investors seeking to buy, if the search process is more efficient, the measure of sellers-to-be "waiting" to sell is reduced ($\frac{\partial \eta_s}{\partial \lambda} < 0$). Proposition 3 presents these results:

Proposition 3. *Given g constant, the measure of buyers-to-be and sellers-to-be is decreasing in* λ ($\frac{\partial \eta_b}{\partial \lambda}, \frac{\partial \eta_s}{\partial \lambda} < 0$) *while the measure of non-searcher investors increases in* λ ($\frac{\partial \eta_0}{\partial \lambda} > 0$)*.*

The proof is in Appendix A.

4.2 Expected Utilities and Price

We now determine the expected utility of the buyers-to-be (v_b), the non-searcher investors (v_0) and the sellers-to-be (v_s), as well as the price p. Investors exit this market because of a need for cash. We assume that the expected utility of outside investors is zero. Once they are out of the market, investors have different investment opportunities and decide where to invest next. They could even choose to re-enter this market again.

To derive the expected utility of every type of investor we analyse the possible transitions between types. For example, a buyer-to-be can leave the market if he receives a liquidity shock, remain a potential buyer or meet a seller-to-be and become a non-searcher. This is summarised in Figure 2:

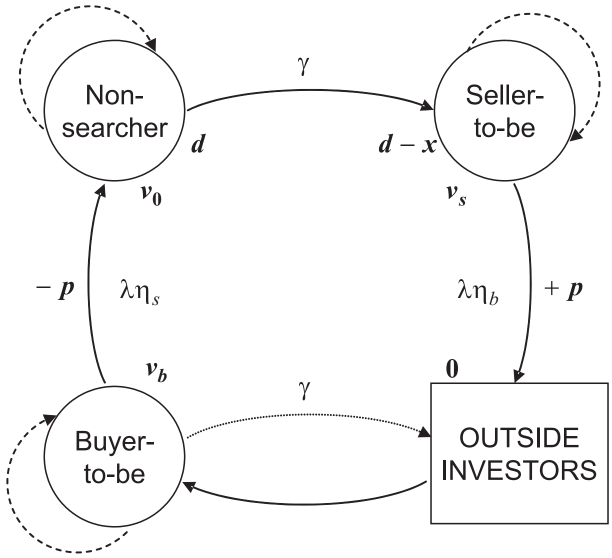


Figure 2: Groups of investors and transitions between groups.

The utility flow rv_b of buyers-to-be is thus equal to the expected flow of exiting the market ($(0 - v_b)\gamma$) and becoming an outside investor plus the expected flow derived from meeting a trading partner seeking to sell (which occurs at rate $\lambda\eta_s$), buying the asset (paying p) and becoming a non-searcher investor ($\lambda\eta_s(v_0 - v_b - p)$). Then,

$$rv_b = -\gamma v_b + \lambda\eta_s(v_0 - v_b - p) \tag{7}$$

Non-searcher investors can either remain non-searchers enjoying the full value d of the asset's dividend flow or receive a liquidity shock with probability γ and become a seller-to-be. In this case, the flow of utility of being a non-searcher is

$$rv_0 = d + \gamma(v_s - v_0) \tag{8}$$

Sellers-to-be exit the market as soon as they meet a trading partner, i.e., with intensity $\lambda\eta_b$ they sell the asset (receiving p) and become outside investors with zero expected utility.

Meanwhile, they enjoy a low level $d - x$ of utility. Thus,

$$rv_s = (d - x) + \lambda\eta_b(p + 0 - v_s) \tag{9}$$

The asset price is determined by bilateral bargaining between a buyer-to-be and a seller-to-be. We have assumed that with probability $\frac{z}{1+z}$ the buyer-to-be makes a take-it-or-leave-it offer to his trading partner and offers him his reservation value v_s. With probability $\frac{1}{1+z}$, the seller-to-be is chosen to offer the buyer-to-be his reservation value $v_0 - v_b$. As a result,

$$p = \frac{z}{1+z}v_s + \frac{1}{1+z}(v_0 - v_b) \tag{10}$$

where z measures the buyer's-to-be bargaining power which we treat as exogenous. Proposition 4 summarises this subsection's main result. The proof is in Appendix A.

Proposition 4. *Given g constant, the system of equations (7)-(10) has a unique solution given by:*

$$\begin{aligned}
v_b &= k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\frac{\lambda\eta_s z}{r+\gamma} && (11)\\
v_0 &= \frac{d}{r} - k\Big(\frac{x}{r} + \frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big)\frac{\gamma}{r+\gamma} && (12)\\
v_s &= \frac{d}{r} - k\Big(\frac{x}{r} + \frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big) && (13)\\
p &= \frac{d}{r} - k\frac{x}{r} && (14)
\end{aligned}$$

where $k = \dfrac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$.

The price of the asset as given by equation (14) is thus equal to the present value of all future dividend flows d, discounted at the rate r, minus a price discount due to illiquidity. The second term is the product of present value of the holding cost x borne by investors seeking to exit the market and a function k. $k \in (0, 1)$ measures the severity or intensity of the illiquidity discount[11].

It is interesting to highlight that the asset price will be higher when fundamentals are

[11] See Section 6 for a discussion of market liquidity.

stronger (i.e. if the asset pays a higher dividend flow d) and whenever the demand for the asset increases ($\frac{\partial p}{\partial d}, \frac{\partial p}{\partial \eta_b} > 0$). On the contrary, the price decreases with investors trying to sell the asset and in the buyer's-to-be bargaining power ($\frac{\partial p}{\partial z}, \frac{\partial p}{\partial \eta_s} < 0$). If during the bargaining process the buyer-to-be holds a more favourable position, he would try to lower the price paid to acquire the asset. The proof of this set of comparative statics is presented in Appendix B.

4.3 Trade among Investors

In this subsection we prove a result we have assumed so far in our analysis:

Proposition 5. *All meetings between buyers-to-be and sellers-to-be result in trade.*

Proof. Trade between buyers-to-be and sellers-to-be occurs if the gain from trade is strictly positive, i.e., if the buyers'-to-be reservation value $v_0 - v_b$ exceeds the sellers'-to-be reservation value v_s. Let us see if $(v_0 - v_b) - v_s > 0$. Subtracting equations (13) and (11) from (12), we get:

$$(v_0 - v_b) - v_s = \frac{x(1+z)}{(r+\gamma)(1+z) + \lambda\eta_s z + \lambda\eta_b}$$

which is always strictly greater than zero since $x, r, \gamma, \lambda, \eta_s, \eta_b, z > 0$.

Therefore, once investors meet, trade among partners always occurs. □

4.4 Thick market externality

In financial markets, thick market externalities arise when the gains from investing in a market depend on the number of investors who decide to come to the market. In this case, the more traders move into a market, the easier become the transactions and as a result the bigger is the gain derived from participating in this market. In our framework, the price of the asset is higher as the flow of investors moving into the market increases[12] ($\frac{\partial p}{\partial g} > 0$). As investors arrive to this market, the costs of search are reduced and hence the illiquidity discount is diminished. This increases the returns to investing in this market, making it more

[12]The proof is presented in Appendix B.

attractive to new investors. To understand how higher participation may encourage further participation we need to endogenise investors' entry decisions.

5 Equilibrium

In our setting, market equilibrium is determined by the fraction of investors entering the market, a measure of each group of investors, their expected utilities and the price of the asset. We centre our study on the steady-state analysis. In the previous section we take as given investors' decision to enter the market, and we now endogeneise the entering rule in Subsection 5.1. A formal definition of the market equilibrium is then presented in Subsection 5.2. Subsection 5.3 introduces the congestion effect.

5.1 Entering Rule

In this subsection we endogenise the entering rule. In our framework, outside investors can choose between entering the market, which we will refer to as *our market*, and investing in an alternative market. Investors are heterogeneous in their outside investment opportunities κ, i.e. each class of investor has access to different investment opportunities. However, once they enter our market, their type no longer influences their decisions in the sense that every buyer-to-be, for instance, enjoys the same expected utility independently of his original outside opportunity. Interestingly, a buyer's-to-be expected utility does depend on the flow of investors who entered this market before him.

Let us refer to the investor who is deciding between moving or not into our market as the marginal investor. And, let us denote by κ' and by $v_{alt}(\kappa')$, respectively, the best outside investment opportunity of the marginal investor and his expected utility from investing in that alternative market. For simplicity, we assume $v_{alt}(\kappa') = \kappa'$, such that an investor with a better outside option (higher κ) enjoys a higher level of expected utility.

When an investor faces the decision to choose a market, he prefers to enter our market if the expected utility v_b of being a buyer-to-be in this market is higher than the expected utility v_{alt} derived from his best outside option. Then, if our market represents the best opportunity for the marginal investor, it is also preferred by any other investor with a worse investment opportunity, i.e. any investor whose type $\kappa < \kappa'$ moves into our market too. As a result, when our market is chosen by a marginal investor with a high type, a high flow of investors enters our market. A high flow of investors implies an increase in the measure of buyers-to-be, which then affects the expected utility of being a buyer-to-be. Thus, even though each investor's type does not alter his expected utility, the type of the last investor who enters does. The type of this last investor defines the total flow who prefers our market and hence determines how concentrated the population of buyers-to-be is.

Let us define the fraction $\nu(\kappa)$ of investors with outside investment opportunity κ who enters the market as follows:

$$\nu(\kappa) = \begin{cases} 0 & \text{if} \quad \kappa > \kappa' \\ [0,1] & \text{if} \quad \kappa = \kappa' \\ 1 & \text{if} \quad \kappa < \kappa' \end{cases}$$

where $1 - \nu(\kappa)$ represents the fraction of investors with outside option κ who invests in alternative markets. The total flow of investors moving into our market is thus given by: $g(\kappa') = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu(\kappa) f(\kappa) d\kappa$, where f defines the total flow of investors entering the economy. In equilibrium, as we discuss in more detail in the next subsection, the total flow g^* depends on the equilibrium fraction of investors ν^* entering our market. But the equilibrium fraction of investors is determined by the marginal investor who is indifferent between our market and his best outside option. We refer to this investor as the indifferent investor. For the indifferent investor, the expected utility of being a buyer-to-be equals the expected utility of his best outside option:

$$v_b \left(g^* = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu^*(\kappa) f(\kappa) d\kappa \right) = v_{alt}(\kappa^*) \tag{15}$$

Before we proceed, let us introduce the formal definition of market equilibrium.

5.2 Equilibrium Definition and Characterisation

Definition 1. *A* ***market equilibrium*** *consists of a fraction $\nu(\kappa)$ of investors entering the market, measures (η_s, η_b, η_0) of investors and expected utilities and prices (v_b, v_0, v_s, p) such that:*

- *(η_s^*, η_b^*, η_0^*) solve the market-clearing condition and inflow-outflow equations given by the system (1) - (3),*
- *(v_b^*, v_0^*, v_s^*, p^*) solve the flow-value equations for the expected utilities and the pricing condition given by the system (7) - (10),*
- *$\nu^*(\kappa)$ solves the entering condition given by the system (15).*

To analyse the equilibria in this market, we need to solve for the fixed points of the system of equations (1) - (3), (7) - (10) and (15). There are two types of possible scenarios depending on the behaviour of the expected utility v_{alt} of investing in an alternative market and the expected utility v_b of being a buyer-to-be in our market. There is an equilibrium where all investors clearly prefer one market (either all enter or no one enters) or an equilibrium where a fraction of investors is better off by investing in our market while others prefer not to enter. Theorem 1 summarises a key result:

Theorem 1. *There is a unique market equilibrium.*

The proof is in Appendix C. To gain some intuition for this result, let us introduce Figure 3. Figure 3 represents the expected utility v_{alt} of investing in an alternative market and the expected utility of being a buyer-to-be of the marginal investor, i.e. the one deciding whether or not to enter our market. Consider, for example, the marginal investor with outside investment opportunity κ_1'. He compares the utility of his outside option, $v_{alt}(\kappa_1') = \kappa_1'$, to the utility of being a buyer-to-be, $v_b(g(\kappa_1'))$, given that investors with outside opportunities $\kappa < \kappa_1'$ have already entered our market. He enters since $v_b(g(\kappa_1')) > v_{alt}(\kappa_1')$, as shown in Figure 3. Now, let us focus on the marginal investor with investment opportunity κ_2'. The expected utility of being a buyer-to-be has decreased because now all investors with $\kappa < \kappa_2'$ are in the market. Still he is better-off by moving into our market. Suppose marginal investor

κ^* is now facing the entry decision. For him, $v_b(g(\kappa^*)) = v_{alt}(\kappa^*)$ and he is indifferent between markets. Any investor with a better outside opportunity prefers not to enter.

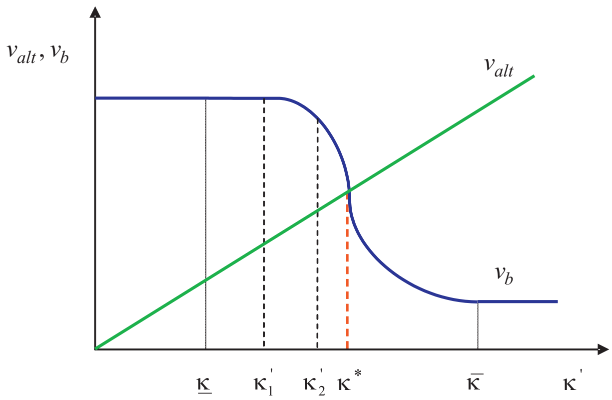


Figure 3: Unique Market Equilibrium - Investors compare expected utilities v_b and v_{alt} and decide to move into our market if $v_b > v_{alt}$. κ^* defines the outside investment opportunity which makes investors indifferent between entering or not our market. $\underline{\kappa}$ and $\overline{\kappa}$ determine the support of the flow of investors who enter the economy.

Let us see why the equilibrium is unique. Given non-negative expected utilities, if $v_b(\kappa' = 0) > v_{alt}(\kappa' = 0)$ and v_b decreases in κ' while v_{alt} is strictly increasing, then by continuity there exists a unique threshold κ^* such that expected utilities are equal and investors indifferent between markets. A unique threshold κ^* then defines a unique flow of investors $g^* \equiv g(\kappa' = \kappa^*)$ entering our market. And given a unique flow of investors g^*, steady-state measures, expected utilities and the asset price can be determined uniquely as stated in Propositions 1 and 4. Consequently, market equilibrium is unique. It is interesting to note that the expected utility of buyers-to-be decreases as more investors enter our market. We discuss this result in the following subsection.

5.3 Market Congestion

Why is the expected utility of a buyer-to-be reduced as the flow of investors entering the market rises? Because buyers-to-be suffer from a *congestion effect*. In our market, an increase in the flow of investors g affects differently the steady-state measures of investors. Every investor who enters our market becomes a buyer-to-be first. Then, only a proportion of buyers-to-be meets a trading partner, purchases the asset and becomes a non-searcher. Only a fraction of non-searcher receives a liquidity shock becoming a seller-to-be. But, given that the measure of buyers-to-be has increased, it is now easier for a seller-to-be to meet a trading partner and hence the steady-state measure of investors seeking to sell is reduced as the flow of investors g increases[13]. As a result, in our framework buyers-to-be are worse off when g rises because it is now more difficult for them to meet a seller-to-be and purchase the asset. There is a congestion effect as investors move into our market in the sense that buyers-to-be face a crowded market where there is increasing competition among buyers-to-be for the fewer sellers-to-be.

6 Liquidity, Market Efficiency and Welfare

In this section we first discuss the relationship between market liquidity and the equilibrium flow of investors who move into our market. In our model, the equilibrium flow is endogenously determined and depends on the characteristics defining the market, the asset and the investors. To analyse, for instance, the consequences on market liquidity of a change in market efficiency we need to understand both the direct effect of this change on the asset price, and hence on liquidity, and also the indirect effect through the equilibrium flow of investors. Subsection 6.2 examines the introduction of a new electronic system in our market to provide some intuition for the interaction between search costs and the equilibrium flow of investors and thus to better understand this indirect effect. The general relationship between the flow of investors and the parameters of the model, including search efficiency, is derived in Subsection 6.3. Finally, in Subsection 6.4 we introduce welfare and study the implications on welfare and market liquidity of an improvement in the efficiency of the search process when our market experiences a fire sale.

6.1 Market Liquidity

In our model, an investor willing to buy or sell needs to find a trading partner and bargain over the asset price before the transaction takes place. Investors cannot trade instantaneously

[13]Comparative statics of the steady-state measures of investors in the market were introduced in Section 4.1 (See Proposition 2).

but there is a time delay due to this search process. This search cost can be identified with the expected time required to locate a trading partner and, as a result, liquidity can be viewed as inversely related to this time delay. In Subsection 4.2 we define illiquidity as measured by the illiquidity discount

$$k\frac{x}{r}$$

where $\frac{x}{r}$ is the present value of the holding cost x and $k = \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$. Let us denote by $\tau^s \equiv \frac{1}{\lambda\eta_b}$ the expected time required to locate a buyer-to-be and by $\tau^b \equiv \frac{1}{\lambda\eta_s}$ the expected time it takes for a buyer-to-be to meet a seller-to-be. The function k is increasing in τ^s and decreasing in τ^b. Then, as the time a seller-to-be needs to wait before he can leave the market (τ^s) increases, k rises and the effect of the illiquidity discount is more severe. In contrast, if a buyer-to-be needs to wait longer to locate a seller-to-be, the effect of illiquidity discount is diminished. The equilibrium level of market liquidity thus rises in η_b^* but diminishes in η_s^*. In our market, an increase in the equilibrium measure of buyers-to-be and a reduction in the equilibrium measure of sellers-to-be occurs whenever the equilibrium flow of investors, g^*, moving into our market increases[14]. We formalise this result in the following proposition, which we prove in Appendix D:

Proposition 6. *Liquidity increases in the flow of investors entering the market.*

Understanding the relationship between market liquidity and investors' decision to enter a market constitutes one of the main motivations of our analysis. In our model, there is a trading externality as the arrival of new investors facilitates the search process for every investor in the market. If the flow of potential traders increases, trade becomes easier and liquidity rises.

However, as the flow of investors moving into a market increases, the congestion effect makes it more difficult for a buyer-to-be to locate a trading partner. Consequently, as the market gets crowded, it becomes less attractive to investors. This translates into a lower flow of investors entering the market and as a result into a less liquid market.

In equilibrium, the flow of investors and hence the level of market liquidity result from a tradeoff between thick market complementarities and congestion effects. The equilibrium

[14]See Proposition 2.

flow of investors though is determined endogenously in our framework and depends on the market, investors and asset characteristics such as search efficiency, frequency of the liquidity shocks and dividend flow, among others. If we were interested in the consequences on market liquidity of a change in any of these characteristics, we would need to consider two different type of implications. Assume, for instance, an improvement in the efficiency of the search process. It would not only increase the rate at which investors meet but it would also affect the flow of investors who enter our market. Specifically, it raises the frequency of meetings between buyers-to-be and sellers-to-be, favouring market liquidity, and induces two opposite effects on the equilibrium flow of investors who move into our market. First, trading externalities attract potential investors, increasing the flow. But, secondly, congestion deters investors from entering our market, diminishing the flow. The overall level of market liquidity thus depends on this tradeoff and on the effect of the improvement in efficiency on the trading frequency. We discuss the aggregate effect on market liquidity in Subsection 6.4, but we first introduce the following example to better understand the interaction between trading costs and flow of investors.

6.2 An Example of a Technological Innovation

We consider a search-based market of asset trading as the one described in the previous sections. We are interested in understanding the consequences of a technological innovation intended to increase the efficiency of the search process, such as the introduction of a new electronic trading system. The efficiency of the search process in our model is defined by the parameter λ. A high value of λ represents an efficient search process and corresponds to a market where the rate at which investors meet trading partners is high and hence the friction introduced by the search process and its associated cost are low.

We assume the flow of investors f entering the economy is uniformly distributed[15] with support $[0, 5]$, where $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$. Investors have time preferences with discount rate equal to 1% ($r = 0.01$). The asset pays a dividend flow $d = 2$ and is in total supply

[15]Formally, we assume a beta distribution defined on the interval $[0, 5]$ with shape parameters $a = 1$ and $b = 1$, which is identical to a uniform distribution with support $[0, 5]$. The beta distribution is a flexible class of distributions defined on the unit interval $[0, 1]$, whose density function may take on different shapes depending on the choice of the two parameters. These include the uniform density function and hump-shaped densities (See Evans et al. (1993)). We introduce the beta distribution to facilitate the comparison between settings when we later discuss the second example.

$S = 2$. The holding cost is defined as a 40% of the dividend flow to indicate that once an investor receives a liquidity shock his valuation of the asset drops to a 60% of the initial value. Liquidity shocks arrive at a Poisson rate $\gamma = 0.2$ and hence the expected time between shocks is 5. The value of z is chosen such that buyers-to-be and sellers-to-be have the same bargaining power, i.e. $z = 1$. We refer to this example as the baseline setting.

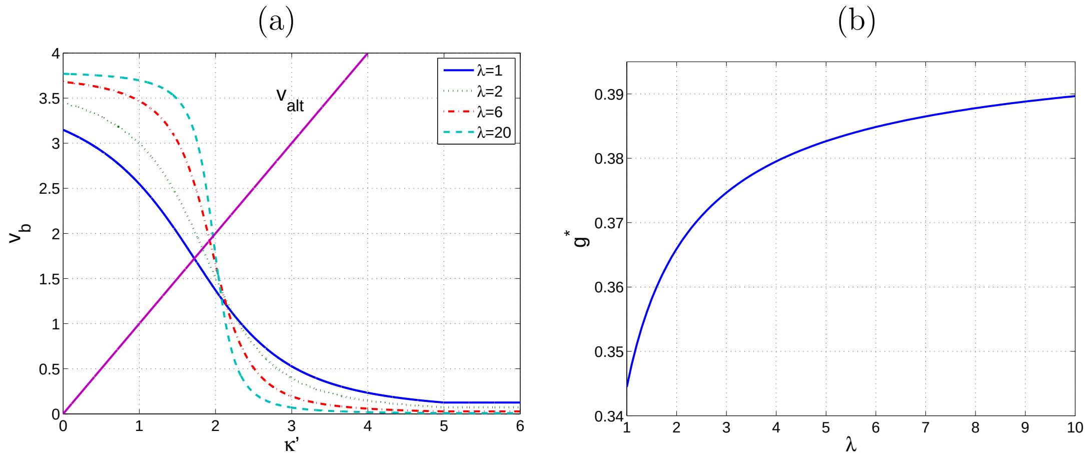


Figure 4: Baseline Setting - Improving efficiency (higher value of λ) attracts more investors to our market (higher g^*). The value of the model parameters is set at the following: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Figure 4(a) represents the expected utility v_{alt} of investing in an alternative market and the expected utility v_b of being a buyer-to-be as a function of the marginal investor's outside investment opportunity κ'. The expected utility v_b of buyers-to-be is plotted for four different values of the market efficiency λ, where a higher λ indicates a more efficient search process. The intersection between v_b and v_{alt} gives, for each level of search efficiency, the threshold κ^* that defines the indifferent investor. κ^* is hence the solution to our fixed point problem. In equilibrium, investors whose best outside investment opportunity κ' is below the threshold value κ^* enter our market, while those with $\kappa' > \kappa^*$ prefer the alternative market. Figure 4(a) shows that an improvement in the efficiency of the search process (higher value of λ) makes our market attractive to more investors (higher κ^*). A higher threshold κ^* then corresponds to an increase in the flow of investors g^* who prefer our market. Figure 4(b) depicts the equilibrium flow of investors g^* entering our market, which is strictly increasing in the efficiency of the search process.

An increase in the equilibrium flow of investors g^* causes a rise in the equilibrium measures of buyers-to-be η_b^* and non-searchers η_0^* and a reduction in the equilibrium measure of sellers-to-be[16] η_s^*. But the equilibrium measures of investors in our market also depend on the efficiency of the search process[17]. In particular, as the search process becomes more efficient (higher value of λ), the measures of investors "waiting" to buy or sell (η_b^* and η_s^*) decrease while the measure of non-searchers rises. The overall effect on the equilibrium measures is presented in the top panel of Figure 5(a):

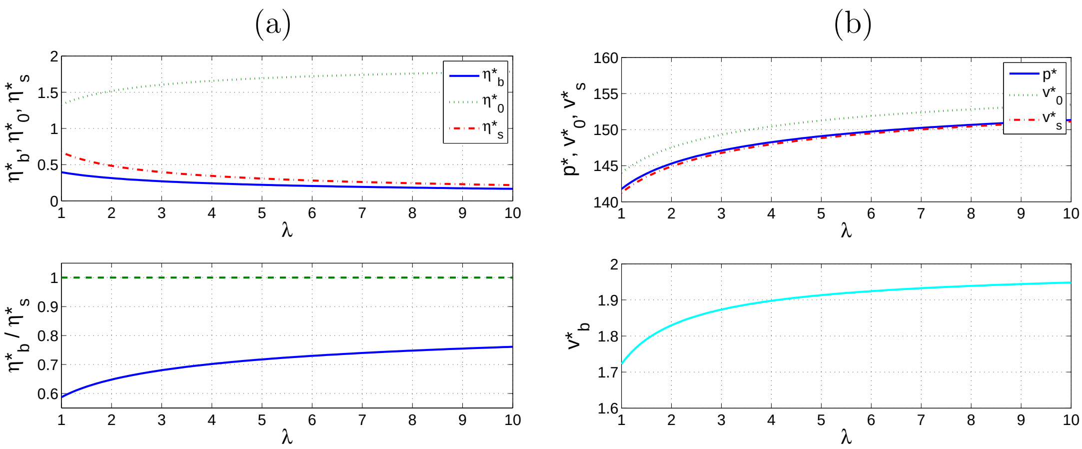


Figure 5: Baseline Setting - Equilibrium measures of investors in our market and ratio of buyers-to-be to sellers-to-be (a), expected utilities and price (b) as a function of the market efficiency λ. Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

More interestingly, the bottom panel of Figure 5(a) illustrates the ratio between buyers-to-be and sellers-to-be as a function of the efficiency of the search process. This ratio captures the notion of congestion in our market. A high value of the proportion of buyers-to-be to sellers-to-be ($>> 1$) describes a market where there is strong competition among buyers-to-be for the few sellers-to-be. There is congestion on the buyer-side in this market. On the contrary, a very low value of this ratio corresponds to a market where there is congestion on the sell-side (more sellers-to-be than buyers-to-be). The effect of congestion gets attenuated as the ratio between buyers-to-be and sellers-to-be tends to 1 as in our baseline setting.

Figure 5(b) depicts equilibrium price and expected utility of sellers-to-be and non-searchers

[16]See Proposition 2.
[17]See Proposition 3.

(top panel) and buyers-to-be (bottom panel) as a function of λ. Price and expected utilities increase in the efficiency of the search process.

In this baseline setting a new electronic trading system, which improves search efficiency, enhances the attractiveness of our market. But this is not always the case. Let us introduce the following example.

6.2.1 Market Boom or Market Crash in an Outside Market

Assume a scenario similar to the one we have just discussed in the baseline setting and let us now consider a severe adverse shock which affects investors' outside investment opportunities. The worsening of investors' outside options could correspond to a boom in our market or to a market crash in another market[18] and would affect the distribution of investors f entering the economy as a function of their outside investment opportunities κ. It would lead to a shift to the left of the mass of the distribution of investors f. In particular, we consider a beta distribution with support $[0, 5]$ and parameters $a = 2$ and $b = 15$, which is a right-skewed hump-shaped density function.

We present our results in Figures 6 and 7, where the value of all parameters (but the distribution parameters) remains as in the baseline setting, i.e., $r = 0.01$, $d = 2$, $S = 2$, $x = 0.4d$, $\gamma = 0.2$ and $z = 1$.

Figure 6(a) illustrates the expected utility v_{alt} of investors' outside options and the expected utility v_b of being a buyer-to-be in our market for the same four values of market efficiency considered in the baseline setting. It is interesting to highlight that the equilibrium threshold κ^* now decreases in the search efficiency, such that to a market with a more efficient search process corresponds a lower cutoff value κ^* of the outside option, which then defines a lower equilibrium flow of investors g^* entering the market. As Figure 6(b) clearly shows, the equilibrium flow of investors entering our market strictly decreases in the search efficiency.

Why is the equilibrium flow of investors decreasing as the search process becomes more efficient? Let us see why this is the case. Our market is now attractive to more investors because of the worsening of conditions in another market. This is indicated in Figures 4(b)

[18]In either case, market conditions improve significantly in our market compared to those in alternative markets. For the ease of exposition, we consider the market crash interpretation.

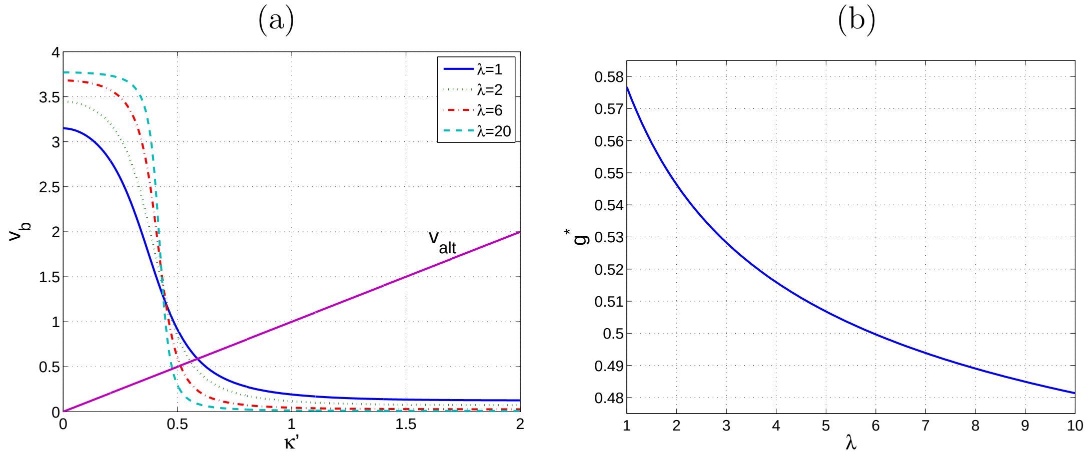


Figure 6: MARKET CRASH SETTING - Improving efficiency (higher value of λ) discourages investors from entering our market (lower equilibrium flow of investors g^*). Model parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

and 6(b), which show that for any given level of market efficiency (fixing λ), the equilibrium flow of investors now entering our market is higher than in the baseline setting. Then, from the buyers'-to-be perspective, our market has become crowded in the sense that there are too many buyers-to-be for each investor seeking to sell and hence it is now more difficult to meet a trading partner and purchase the asset. If search frictions were then reduced in this market (higher values of λ), the effect of congestion would be amplified. Investors would meet at faster rates, which reduces the measures of buyers-to-be and sellers-to-be as shown in the top panel of Figure 7(a) but, most importantly, it would allow sellers-to-be to exit faster leading to an even more unbalanced distribution of investors (bottom panel of Figure 7(a)).

As the bottom panel of Figure 7(b) illustrates, buyers-to-be are worse-off as the search process becomes more efficient and congestion intensifies. This discourages potential investors from moving into our market, reducing the equilibrium flow of investors g^*.

The reason for this counterintuitive result is that lower trading frictions in a one-sided market magnify the effect of congestion, discouraging investors from entering this market. In this case, congestion dominates thick market externalities and hence the introduction of a measure intended to improve market efficiency results in a less attractive market.

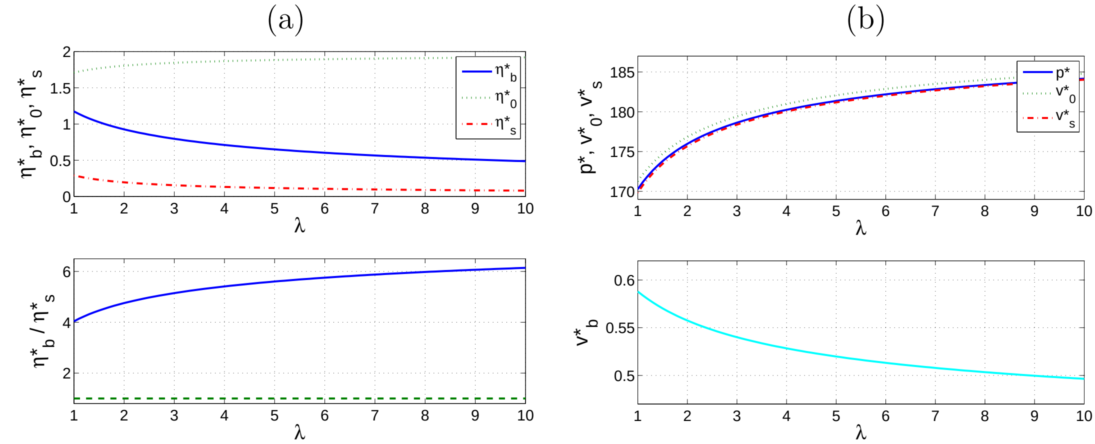


Figure 7: Market Crash Setting - Equilibrium measures of investors in our market and ratio of buyers-to-be to sellers-to-be (a), expected utilities and price (b) as a function of the market efficiency λ. Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

6.3 Flow of Investors and Market Efficiency

In this subsection we determine the general relationship between the equilibrium flow of investors g^* entering the market and the efficiency λ of the search process. To simplify the analysis we first derive the equilibrium measure of sellers-to-be (η_s^*) as a function of the market efficiency λ and the other nine parameters of the model (γ, r, S, x, z, a, b, $\underline{\kappa}$ and $\overline{\kappa}$). There is a one-to-one relationship between g^* and η_s^*. Hence, once we compute η_s^*, we can then determine the equilibrium flow of investors g^* who enter our market.

In our setting, market equilibrium is the solution to the system of equations (1)-(3), (7)-(10) and (15). We thus need to solve for the fixed point of this system, which is reduced to solving the indifference condition that defines investors' entry rule. Investors, in our framework, compare the expected utility v_{alt} of investing in an alternative market to the expected utility v_b derived from being a buyer-to-be in our market and they decide to move in whenever $v_b > \kappa' \equiv v_{alt}$. To present this indifference condition ($v_b = \kappa'$) as a function of the measure of sellers-to-be (η_s), let us first redefine the measure of buyers-to-be η_b as a function of η_s. Using equations (5) and (6) we find:

$$\eta_b = \frac{\gamma}{\lambda} \frac{A}{2\gamma S - A} = \frac{\gamma}{\lambda} \frac{2\gamma(S - \eta_s)}{2\gamma S - 2\gamma(S - \eta_s)} \qquad \Rightarrow \qquad \eta_b = \frac{\gamma}{\lambda} \frac{S - \eta_s}{\eta_s} \tag{16}$$

We can now express the expected utility v_b of buyers-to-be as a function of η_s by substituting equation (16) into equation (11):

$$v_b = \frac{x}{r+\gamma} \frac{\lambda z \eta_s^2}{\lambda z \eta_s^2 + \left[(r+\gamma)(1+z) - \gamma\right] \eta_s + \gamma S} \tag{17}$$

Next we write κ' as a function of η_s. In this model, the flow of investors g who move into our market is determined by the proportion of the total flow of investors f whose expected utility v_b of being a buyer-to-be exceeds their best outside option κ'. We assume the flow of investors f follows a beta distribution with support $[\underline{\kappa}, \overline{\kappa}]$ and shape parameters[19] a and b. For notational convenience we omit reference to the shape parameters. Then,

$$g(\kappa') = \int_{\underline{\kappa}}^{\kappa'} f_{beta}(\kappa) d\kappa = F_{beta}(\kappa') \qquad \Rightarrow \qquad \kappa' = F_{beta}^{-1}(g) \tag{18}$$

where f_{beta} and F_{beta} denote respectively the probability density function (*pdf*) and the cumulative distribution function (*cdf*) of a beta distribution. F_{beta}^{-1} is the inverse cumulative distribution function. Using equation (5) and the definition of A in Page 11 we can express the flow of investors g as a function of the measure of sellers-to-be η_s:

$$g = \gamma \left(1 + \frac{\gamma}{\lambda \eta_s}\right) (S - \eta_s) \tag{19}$$

Substituting equation (19) in equation (18) yields:

$$\kappa' = F_{beta}^{-1}\left(\gamma \left(1 + \frac{\gamma}{\lambda \eta_s}\right) (S - \eta_s)\right) \tag{20}$$

[19] The probability density function of the beta distribution defined over the interval $[0, 1]$ with shape parameters a and b is:

$$f_{beta}(y; a, b) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} y^{a-1} (1-y)^{b-1}$$

where $a, b > 0$ and $\Gamma(\cdot)$ is the gamma function. For integer values of a and b, the cumulative distribution function of the beta distribution is given by:

$$F_{beta}(y; a, b) = \sum_{j=a}^{a+b-1} \binom{a+b-1}{j} y^j (1-y)^{a+b-1-j}$$

where $\binom{a+b-1}{j} = \frac{(a+b-1)!}{j!(a+b-1-j)!}$.

The indifference condition results from equating the expected utility v_b of buyers-to-be (equation (17)) to the marginal investor outside option κ' (equation (20)):

$$\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}=F_{beta}^{-1}\left(\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)\right)$$

Rearranging, we get

$$\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)=F_{beta}\left(\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}\right)$$

Then,

$$\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)=\sum_{j=a}^{a+b-1}\binom{a+b-1}{j}\Big(\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}\Big)^j \Big(1-\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}\Big)^{a+b-1-j} \quad (21)$$

Equation (21) is a polynomial of degree $2(a+b)$ in the measure of sellers-to-be[20]. To solve for η_s^* we use the bisection method[21]. Once we compute η_s^*, we can derive g^*:

$$g^*=g^*(\lambda,\gamma,r,S,x,z,a,b,\underline{\kappa},\overline{\kappa})$$

[20]In the simple case of shape parameters of the beta distribution both equal to 1 ($a=1=b$), which corresponds to a uniform distribution with support $[\underline{\kappa},\overline{\kappa}]$, the indifference condition ($v_b=\kappa'$) is:

$$\frac{x}{r+\gamma}\frac{\lambda z\eta_s^2}{\lambda z\eta_s^2+[(r+\gamma)(1+z)-\gamma]\eta_s+\gamma S}=\underline{\kappa}+(\overline{\kappa}-\underline{\kappa})\gamma\left(1+\frac{\gamma}{\lambda\eta_s}\right)(S-\eta_s)$$

Reorganising terms yields the following polynomial of degree four in the measure of sellers-to-be η_s:

$$\lambda^2 z(\overline{\kappa}-\underline{\kappa})\gamma\eta_s^4 + \Big[\lambda(\overline{\kappa}-\underline{\kappa})\gamma C-\lambda zD+\frac{x}{r+\gamma}\lambda^2 z\Big]\eta_s^3+\Big[\lambda(\overline{\kappa}-\underline{\kappa})\gamma^2 S(1-z)-CD\Big]\eta_s^2 - $$
$$- \Big[\gamma SD+(\overline{\kappa}-\underline{\kappa})\gamma^2 SC\Big]\eta_s-(\overline{\kappa}-\underline{\kappa})\gamma^3 S^2=0$$

where $C=(r+\gamma)(1+z)-\gamma$ and $D=\lambda\underline{\kappa}+\lambda(\overline{\kappa}-\underline{\kappa})\gamma S-(\overline{\kappa}-\underline{\kappa})\gamma^2$. There exists closed-form solution to this equation. In particular, there are at most four solutions but only one, η_s^*, (as proved in Subsection 5.2) lies in the interval $(0,S)$, the set of possible values of the measure of sellers-to-be. Unfortunately, the solution is intractable. We use the bisection method over the interval $[0,S]$ to determine the zero of this equation.

[21]The bisection algorithm is a numerical method for finding the root of a function. It recursively divides an interval in half and selects the subinterval containing the root, until the interval is sufficiently small. Burden and Faires (1993) presents a clear description of this algorithm as well as other numerical methods for solving root-finding problems.

where g^* is a function of the efficiency of the search process λ, the rate γ at which investors receive liquidity shocks, the discount rate r, the supply of the asset S, the holding cost x, buyer's-to-be bargaining power z, the shape parameters a and b of the beta distribution and the support $[\underline{\kappa}, \overline{\kappa}]$ of the flow of investors f entering the economy. To gain some intuition for how the model parameters affect the equilibrium flow of investors g^*, we set the value of those defining the distribution of f and vary the other parameters of the model. We assume $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$ as in the market crash setting in Subsection 6.2.1. The first set of results is depicted in Figure 8:

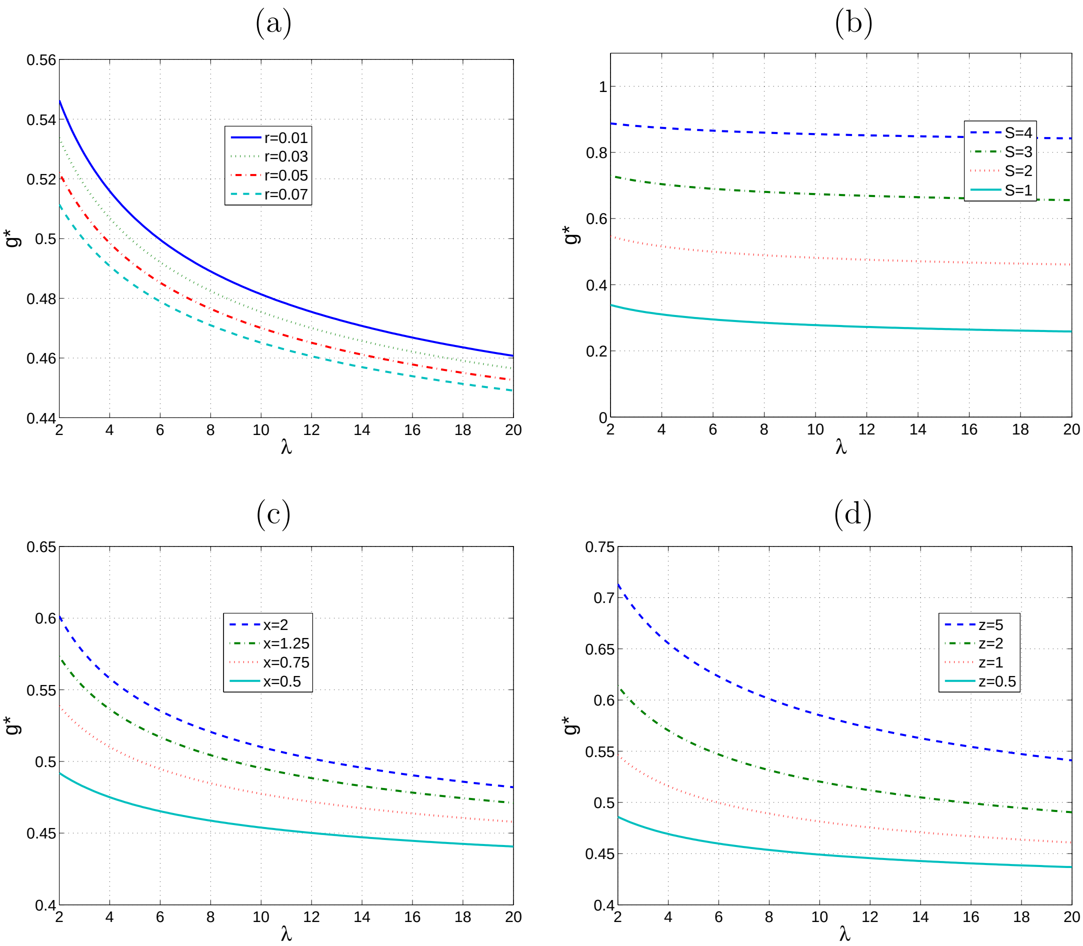


Figure 8: Equilibrium flow of investors g^* entering our market as a function of the market efficiency λ for different values of r (a), S (b), x (c) and z (d). Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.2$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Figure 8 represents g^* as a function of the efficiency of the search process λ, where g^*

is plotted for four different values of the discount rate r (a), the supply of the asset S (b), the holding cost x (c) and the buyers'-to-be bargaining power z (d). The distribution of parameters underlying these graphs corresponds to a one-sided market scenario discussed in Subsection 6.2.1. Then, in all four cases, increasing market efficiency (higher values of λ) translates into a lower equilibrium flow of investors entering the market. Also, for a given level of market efficiency (fixed λ), more investors move into our market as we increase the total supply of the asset, the holding cost or the buyers'-to-be bargaining power. The equilibrium flow of investors decreases as they become more impatient (higher r).

More interesting is the interaction between market efficiency λ and the arrival rate of liquidity shocks γ. Figure 9(a) demonstrates how the equilibrium flow of investors g^*, who enter our market, varies with the market efficiency λ and the frequency of liquidity shocks γ. Contours are depicted in Figure 9(b). Now, the relationship between g^* and λ is non-monotonic. It is first decreasing in market efficiency, corresponding to a one-sided market scenario, but then it becomes increasing in λ for higher values of the liquidity shock rate γ.

If liquidity shocks arrive at very low rates (low values of γ), investors hold the asset, on average, for a long time. As a result, there are few investors trying to sell and exit the market. Increasing the efficiency of this market (raising λ) attracts new investors, amplifying the effect of congestion. The market becomes one-sided because there are more buyers-to-be and few sellers-to-be. In this case, reducing market frictions diminishes the flow of investors. This is shown in Figure 9(c) for values of $\gamma \leq 0.3$. This phenomenon is attenuated as investors need to exit at a faster rate. Then, for intermediate values of γ, there are enough sellers-to-be in our market and improving market efficiency attracts new investors ($\gamma = 0.4$ and $\gamma = 0.5$ in Figure 9(c)). Thick market externalities dominate congestion. Also, as Figure 9(d) indicates, if investors need for cash is very frequent (values of γ above 0.5), they prefer not to invest and the flow of investors g^* who enter our market is reduced. Still, for a given frequency of the liquidity shocks γ, diminishing search frictions improves the attractiveness of our market.

6.4 Liquidity and Welfare

In this subsection we discuss market liquidity and present the welfare analysis. In particular, we are interested in the implications of potential policies designed to improve the efficiency

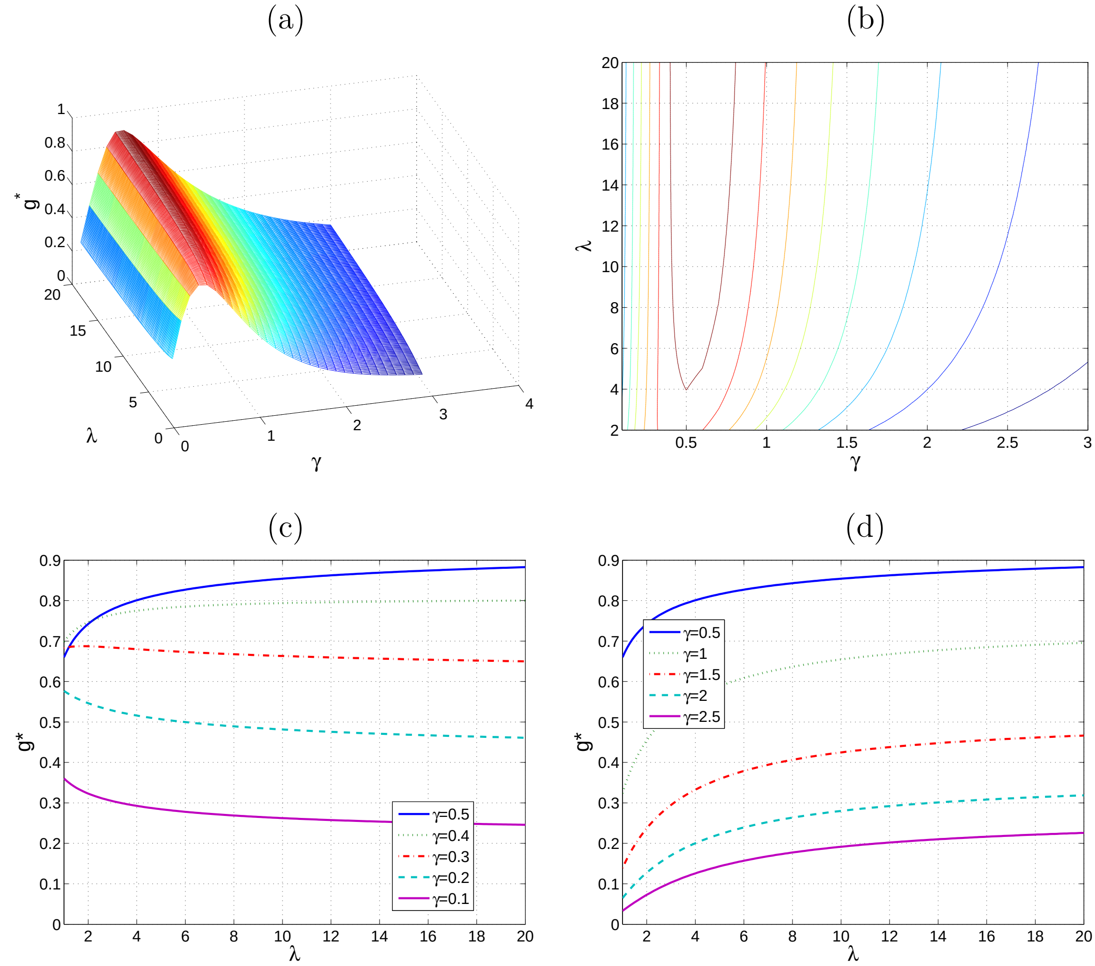


Figure 9: Equilibrium flow of investors g^* entering our market as a function of the market efficiency λ and the frequency of liquidity shocks γ. The values of other parameters of the model are set at the following: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $a = 2$, $b = 15$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

of the search process and to thus reduce market frictions.

We measure welfare by the weighted sum of investors' expected utilities. Weights are determined by the measure of every type of investors in our economy, including the outside investors. Then, our measure of welfare is:

$$W = \eta_b v_b + \eta_0 v_0 + \eta_s v_s + \int_{\kappa^*}^{\overline{\kappa}} \kappa f(\kappa) d\kappa \tag{22}$$

where the first three terms represent the welfare of the investors who prefer to enter our market ($W_{\text{inside investors}}$) and the last term reflects the welfare of outside investors ($W_{\text{outside investors}}$).

Outside investors (those with investment opportunities above the threshold value κ^*) enjoy the expected utility derived from investing in an alternative market v_{alt}, which for simplicity we assume equal to κ', the outside investment opportunity. Substituting equations (11)-(13) and the *pdf* of a beta distribution into equation (22), we get:

$$W_{\text{inside investors}} = \frac{d}{r}S - kx\frac{1}{r+\gamma}\frac{\frac{\gamma}{r}S\big[(\gamma+\lambda\eta_s)z+(r+\gamma)\big]+\eta_s\big[(r+2\gamma+\lambda\eta_s)z+(r+\gamma)\big]}{(r+\gamma+\lambda\eta_s)z+\gamma}$$

$$W_{\text{outside investors}} = \frac{a}{a+b}\Big[1-F_{beta}(\kappa^*;a+1,b)\Big] = \frac{a}{a+b}\left[1-\sum_{j=a+1}^{a+b}\binom{a+b}{j}(\kappa^*)^j(1-\kappa^*)^{a+b-j}\right]$$

To gain some intuition for how changes in market efficiency affect welfare we introduce the last example.

6.4.1 Fire Sales in our Market

Consider a search-based market similar to the baseline setting described in Subsection 6.2 and assume investors need for cash is now more frequent. Specifically, we assume liquidity shocks arrive at a Poisson rate $\gamma = 0.4$. The value of all other parameters remains as in the baseline case: $r = 0.01$, $d = 2$, $S = 2$, $x = 0.4d$, $z = 1$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Investors now prefer to hold the asset, on average, for a shorter period of time and they are willing to sell and exit our market at faster rates. Then, for any given value of the search efficiency, the equilibrium measure of sellers-to-be has increased significantly (top panel of Figure 10(a)) compared to the baseline setting (top panel of Figure 5(a)), while the equilibrium measure of non-searchers has decreased. Given that there are now more sellers-to-be in our market, it is easier for an investor seeking to purchase the asset to meet a trading partner. As a result, the equilibrium measure of buyers-to-be has diminished compared to the baseline case. Most importantly, the proportion of buyers-to-be to sellers-to-be has fallen drastically. This is depicted in the bottom panel of Figure 10(a). Our market is now one-sided and there is severe congestion on the sell-side of the market. This scenario could correspond to a market experiencing a fire sale.

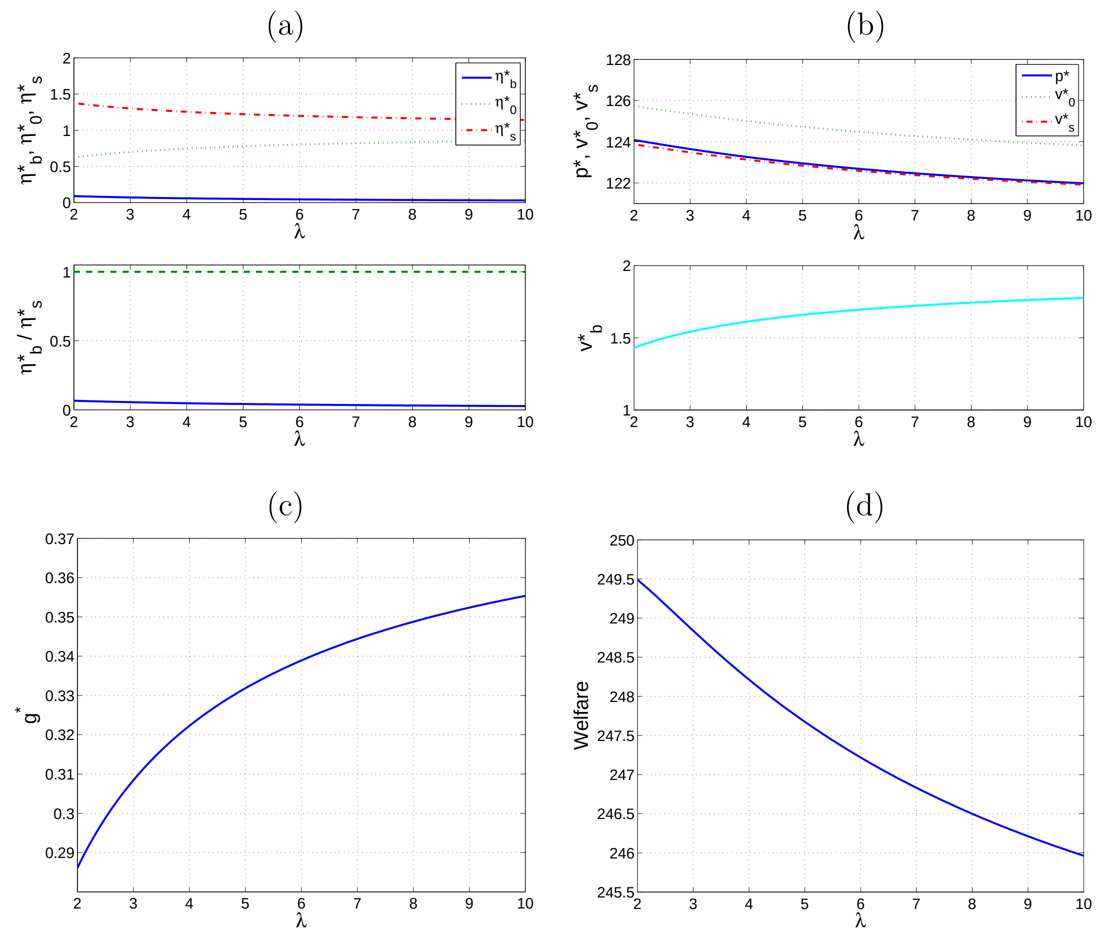


Figure 10: Fire Sales Setting - Equilibrium measures of investors in our market and ratio of buyers-to-be to sellers-to-be (a), expected utilities and price (b), flow of investors g^* (c) and welfare (d) as a function of the market efficiency λ. Other parameters are set at the following values: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.4$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Increasing the efficiency of the search process (higher value of λ) in this market causes two effects. First, it raises the flow of investors who enter our market as plotted in Figure 10(c). Secondly, investors meet at faster rates reducing the equilibrium measure of buyers-to-be and sellers-to-be (top panel of Figure 10(a)). The overall effect on the ratio of buyers-to-be to sellers-to-be is presented in the bottom panel of Figure 10(a). As the market becomes more efficient, the proportion of buyers-to-be to sellers-to-be falls further and from the sellers'-to-be perspective the market gets even more crowded. Congestion intensifies as it is now more difficult to meet a buyer-to-be and exit the market. Hence, as the top panel of Figure 10(b) illustrates, sellers-to-be and non-searchers (who become sellers-to-be at rate γ) are worse-off as efficiency rises. The expected utility of buyers-to-be increases in λ because they can now acquire the asset at faster rates (bottom panel of Figure 10(b)).

A very interesting result is presented in Figure 10(d). We find that as search frictions are reduced, welfare decreases. In this market, improving the efficiency of the search process amplifies the effect of congestion. There are then fewer buyers-to-be per each seller-to-be and the expected utilities of investors holding the asset fall. This induces an adverse effect on welfare.

Figure 11(a) represents our measure of illiquidity as a function of the efficiency of the search process λ, where illiquidity is defined as the price discount. As market efficiency increases and the population of investors gets saturated with sellers-to-be, the price of the asset falls as shown in the top panel of Figure 10(b). This leads to the rise in illiquidity depicted in Figure 11(a). Intuitively, given that there are few buyers-to-be compared to sellers-to-be, the price of the asset behaves as if buyers-to-be would hold a more favourable position in the bargaining process. The effect is equivalent to an increase in the buyers'-to-be bargaining power z, which is exogenous in our model. If we were to endogenise z, the effect on the price (and hence on market liquidity) would be amplified.

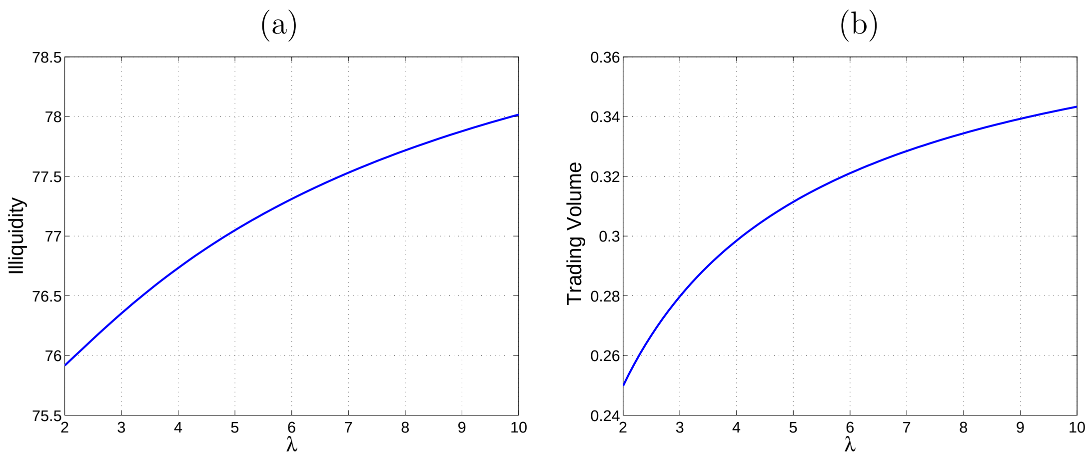


Figure 11: Fire Sales Setting - Illiquidity measured by price discount (a) and trading volume (b) as a function of the market efficiency λ. The value of the model parameters is set at the following: $r = 0.01$, $S = 2$, $x = 0.8$, $z = 1$, $\gamma = 0.4$, $a = 1$, $b = 1$, $\underline{\kappa} = 0$ and $\overline{\kappa} = 5$.

Our market becomes less liquid as the search frictions are reduced. However, as Figure

11(b) indicates, trading volume increases. The reason for this counterintuitive result is the following. Facilitating search in our market has two consequences. First, it magnifies the effect of congestion. There are fewer buyers-to-be relative to the measure of investors trying to exit. Buyers-to-be prefer to pay less to purchase the asset, which translates into a lower price and hence into a less liquid market (higher price discount). Second, it raises the frequency of meeting between trading partners. Investors in our market now meet at a faster rate, increasing the trading volume. Consequently, even though our market is less liquid, investors meet faster and trading volume increases.

7 Conclusions

This paper proposes a search-based model to study the relationship between market liquidity and the endogenous arrival of potential investors to a specific market. As investors enter a market, they make trade easier, attracting new investors. This gives rise to a thick market externality. Interestingly, as investors get attracted to this market, the market becomes crowded and congestion reduces the returns to investing. This paper aims to complement the literature on self-fulfilling liquidity by incorporating a second effect: the congestion effect.

In our market traders can invest in one asset which can be traded only when a pair of investors meet and bargain over the terms of trade. Finding a trading partner takes time and introduces opportunity and other costs. Investors' ability to trade thus affects the illiquidity discount and ultimately, the equilibrium price. We present a numerical example of an advance in trading technology to illustrate the link between the flow of new investors and market liquidity, and to discuss the implications of search frictions on market liquidity.

We then derive the general relationship between the equilibrium flow of investors moving into a market and the efficiency of the search process and highlight the tradeoff between the thick market complementarity and the congestion effect. The equilibrium outcome depends on which of these two effects dominates. In particular, we find that diminishing trade frictions in a market with many buyers and too few sellers leads to a lower equilibrium flow of investors into this market. Less search frictions would allow sellers to exit faster amplifying the effect of congestion (even more buyers per seller) and further discouraging investors from entering this market. We also show that reducing market frictions, in a "congested" market experiencing a fire sale, induces an adverse effect on both market liquidity and welfare. Improving search efficiency (to facilitate coordination and enhance liquidity), magnifies the effect of congestion (less buyers per seller trying to exit) to the detriment of the overall level of market liquidity and social welfare. From this perspective, this paper presents an example of the Theory of the Second Best, where eliminating one but not all market imperfections does not necessary increase efficiency as it may amplify the effect of the remaining distortions.

Appendix

A Proofs of Propositions 1 - 4

Proof of Proposition 1

Proof. Rearranging equation (3), we get

$$h(\eta_0) \equiv \gamma\eta_0^2 - \left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)\eta_0 + Sg = 0$$

where $\eta_0 \in \mathbb{R}_+$. This quadratic function takes positive values as $\eta_0 \to \infty$, is non-negative at $\eta_0 = 0$ and negative at $\eta_0 = S$. Then, by continuity, the polynomial equation has a root in the interval $[0, S)$ and another one in the interval (S, ∞). The two solutions $\eta_0^{(1)}$ and $\eta_0^{(2)}$ are given by:

$$\eta_0^{(1)} = \frac{1}{2\gamma}\Big[\big(g + \gamma S + \frac{\gamma^2}{\lambda}\big) - \sqrt{\big(g + \gamma S + \frac{\gamma^2}{\lambda}\big)^2 - 4\gamma g S}\Big]$$

$$\eta_0^{(2)} = \frac{1}{2\gamma}\Big[\big(g + \gamma S + \frac{\gamma^2}{\lambda}\big) + \sqrt{\big(g + \gamma S + \frac{\gamma^2}{\lambda}\big)^2 - 4\gamma g S}\Big]$$

where $0 \leq \eta_0^{(1)} < S < \eta_0^{(2)} < \infty$. $\eta_0^{(2)}$ is thus not a valid solution since the total supply of the asset is held either by the non-searchers or by the sellers-to-be and as a result the measure of non-searchers cannot exceed the supply of the asset. Then, there is unique solution to equation (3) given by:

$$\eta_0 = \frac{1}{2\gamma}A \tag{A.1}$$

where $A = (g+\gamma S+\frac{\gamma^2}{\lambda}) - \sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2 - 4\gamma g S}$. Plugging equation (A.1) into equations (1) and (2), we find

$$\begin{aligned}
\eta_s &= S - \frac{1}{2\gamma}A \\
\eta_b &= \frac{\gamma}{\lambda}\frac{A}{2\gamma S - A}
\end{aligned}$$

which proves Proposition 1. □

Proof of Proposition 2

Proof. Let us compute the partial derivatives of the measures given by the system of equations (4) - (6) with respect to g:

$$\frac{\partial \eta_0}{\partial g} = \frac{\partial \eta_0}{\partial A}\frac{\partial A}{\partial g} = \frac{1}{2\gamma}\frac{\partial A}{\partial g} \tag{A.2}$$

$$\frac{\partial \eta_s}{\partial g} = \frac{\partial \eta_s}{\partial A}\frac{\partial A}{\partial g} = -\frac{1}{2\gamma}\frac{\partial A}{\partial g} \tag{A.3}$$

$$\frac{\partial \eta_b}{\partial g} = \frac{\partial \eta_b}{\partial A}\frac{\partial A}{\partial g} = \frac{2\gamma^2}{\lambda}\frac{S}{\left(2\gamma S - A\right)^2}\frac{\partial A}{\partial g} \tag{A.4}$$

where

$$\frac{\partial A}{\partial g} = 1 - \frac{(g+\gamma S+\frac{\gamma^2}{\lambda}) - 2\gamma S}{\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} \tag{A.5}$$

To determine the sign of $\frac{\partial A}{\partial g}$, we check if the second term on the right-hand-side of equation (A.5) is greater than 1:

$$\frac{\left(g+\gamma S+\frac{\gamma^2}{\lambda}\right) - 2\gamma S}{\sqrt{\left(g+\gamma S+\frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S}} > 1 \quad ;$$

$$\left(g+\gamma S+\frac{\gamma^2}{\lambda}\right) - 2\gamma S > \sqrt{\left(g+\gamma S+\frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S} \quad ; \tag{A.6}$$

where the right-hand-side of equation (A.6) is strictly positive since

$$\sqrt{(g+\gamma S+\frac{\gamma^2}{\lambda})^2-4\gamma g S}=\sqrt{(g-\gamma S)^2+2(g+\gamma S)\frac{\gamma^2}{\lambda}+\frac{\gamma^4}{\lambda^2}}>0 \tag{A.7}$$

We analyse two cases. If $(g+\gamma S+\frac{\gamma^2}{\lambda})-2\gamma S\leq 0$, then equation (A.6) is not satisfied. On the contrary, if $(g+\gamma S+\frac{\gamma^2}{\lambda})-2\gamma S>0$,

$$\begin{aligned}\Big[\big(g+\gamma S+\frac{\gamma^2}{\lambda}\big)-2\gamma S\Big]^2 &> \Big[\sqrt{\big(g+\gamma S+\frac{\gamma^2}{\lambda}\big)^2-4\gamma g S}\Big]^2;\\ \big(g-\gamma S+\frac{\gamma^2}{\lambda}\big)^2 &> \big(g+\gamma S+\frac{\gamma^2}{\lambda}\big)^2-4\gamma g S;\end{aligned}$$

Simplifying we arrive to:

$$4\frac{\gamma^3}{\lambda}S<0$$

a contradiction, since γ,λ and $S>0$. Therefore, the second term in equation (A.5) is strictly lower than 1 and as a result:

$$\frac{\partial A}{\partial g}>0 \tag{A.8}$$

Thus, substituting the previous equation into equations (A.2) - (A.4) yields:

$$\begin{aligned}\frac{\partial \eta_0}{\partial g} &> 0\\ \frac{\partial \eta_s}{\partial g} &< 0\\ \frac{\partial \eta_b}{\partial g} &> 0\end{aligned}$$

since γ,λ and $S>0$. □

Proof of Proposition 3

Proof. Using equations (4) - (6) we can compute the partial derivatives of the measures of every type of investor with respect to the efficiency of the search process λ:

$$\frac{\partial \eta_0}{\partial \lambda} = \frac{\partial \eta_0}{\partial A}\frac{\partial A}{\partial \lambda} = \frac{1}{2\gamma}\frac{\partial A}{\partial \lambda} \tag{A.9}$$

$$\frac{\partial \eta_s}{\partial \lambda} = \frac{\partial \eta_s}{\partial A}\frac{\partial A}{\partial \lambda} = -\frac{1}{2\gamma}\frac{\partial A}{\partial \lambda} \tag{A.10}$$

$$\frac{\partial \eta_b}{\partial \lambda} = \frac{\gamma}{\lambda}\frac{1}{2\gamma S - A}\left[\frac{2\gamma S}{2\gamma S - A}\frac{\partial A}{\partial \lambda} - \frac{1}{\lambda}A\right] \tag{A.11}$$

where

$$\frac{\partial A}{\partial \lambda} = -\frac{\gamma^2}{\lambda^2}\left(1 - \frac{g + \gamma S + \frac{\gamma^2}{\lambda}}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}}\right) \tag{A.12}$$

We verify whether the second term in the expression in parenthesis is greater than 1 to determine the sign of $\frac{\partial A}{\partial \lambda}$,

$$\begin{aligned} \frac{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)}{\sqrt{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S}} &> 1; \\ \left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 &> \left[\sqrt{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S}\right]^2; \\ \left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 &> \left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S; \end{aligned} \tag{A.13}$$

where we can square both sides of the expression because, using equation (A.7) and g, γ, S and $\lambda > 0$, the numerator and denominator are strictly positive. Rearranging equation (A.13) we get:

$$4\gamma g S > 0$$

which is true since γ, g and $S > 0$. As a result, the second term in the expression in parenthesis in equation (A.12) is strictly greater than 1 and

$$\frac{\partial A}{\partial \lambda} > 0 \tag{A.14}$$

Thus, plugging the previous equation into equations (A.9) - (A.10) we find:

$$\frac{\partial \eta_0}{\partial \lambda} > 0$$
$$\frac{\partial \eta_s}{\partial \lambda} < 0$$

The proof that $\frac{\partial \eta_b}{\partial \lambda} < 0$ is not so straightforward. Let us first rearrange equation (A.12) as follows

$$\frac{\partial A}{\partial \lambda} = \frac{\gamma^2}{\lambda^2} \frac{A}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} \tag{A.15}$$

Now, substituting equation (A.15) in equation (A.11) we get:

$$\frac{\partial \eta_b}{\partial \lambda} = \frac{\gamma}{\lambda^2} \frac{A}{2\gamma S - A} \left[\frac{2\gamma S}{2\gamma S - A} \frac{\gamma^2}{\lambda} \frac{1}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} - 1 \right] \tag{A.16}$$

where we need to derive the sign of the expression in brackets to determine the sign of $\frac{\partial \eta_b}{\partial \lambda}$. Let us then verify if the first term of the expression in brackets in equation (A.16) is strictly lower than 1:

$$\frac{2\gamma S}{2\gamma S - A} \frac{\gamma^2}{\lambda} \frac{1}{\sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}} < 1;$$
$$(2\gamma S - A)\lambda \sqrt{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S} - 2\gamma^3 S > 0;$$
$$\lambda \Big[(2\gamma S - A) \sqrt{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S} - \frac{2\gamma^3 S}{\lambda} \Big] > 0;$$

Given that $\lambda > 0$ and $A = (g + \gamma S + \frac{\gamma^2}{\lambda}) - \sqrt{(g + \gamma S + \frac{\gamma^2}{\lambda})^2 - 4\gamma g S}$, then

$$\Big[2\gamma S - \big(g + \gamma S + \frac{\gamma^2}{\lambda}\big) \Big] \sqrt{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S} + \Big[\sqrt{\left(g + \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - 4\gamma g S} \Big]^2 - \frac{2\gamma^3 S}{\lambda} > 0;$$
$$\left(g - \gamma S + \frac{\gamma^2}{\lambda}\right)^2 - \left(g - \gamma S + \frac{\gamma^2}{\lambda}\right) \sqrt{\left(g - \gamma S + \frac{\gamma^2}{\lambda}\right)^2 + \frac{4\gamma^3 S}{\lambda}} + \frac{2\gamma^3 S}{\lambda} > 0;$$

To simplify the exposition of the proof, let us define $D \equiv g - \gamma S + \frac{\gamma^2}{\lambda}$. Therefore,

$$D^2 - D\sqrt{D^2 + \frac{4\gamma^3 S}{\lambda}} + \frac{2\gamma^3 S}{\lambda} > 0 \tag{A.17}$$

We consider two possible scenarios. If $D \leq 0$, then equation (A.17) is satisfied since λ, γ and $S > 0$. On the contrary, if $D > 0$, then we need to prove that

$$D^2 + \frac{2\gamma^3 S}{\lambda} > D\sqrt{D^2 + \frac{4\gamma^3 S}{\lambda}}$$

Squaring both sides and rearranging, we find

$$D^4 + \frac{4\gamma^3 S}{\lambda}D^2 + \frac{4\gamma^6 S^2}{\lambda^2} > D^2\left(D^2 + \frac{4\gamma^3 S}{\lambda}\right)$$

Simplifying,

$$\frac{4\gamma^6 S^2}{\lambda^2} > 0$$

and this is always satisfied. Then, we have shown that the first term in the expression in brackets in equation (A.16) is strictly lower than 1 and as a result

$$\frac{\partial \eta_b}{\partial \lambda} < 0$$

which completes the proof of Proposition 3. □

Proof of Proposition 4

Proof. Using equation (10), we can rewrite equations (7) and (9) as:

$$rv_b = -\gamma v_b + \lambda \eta_s \frac{z}{1+z}(v_0 - v_b - v_s) \tag{A.18}$$

$$rv_s = d - x + \lambda \eta_b \frac{1}{1+z}(v_0 - v_b - v_s) \tag{A.19}$$

Subtracting equation (A.18) from equation (8) yields:

$$\begin{aligned} r(v_0 - v_b) &= d + \gamma(v_s - v_0) - \Big[-\gamma v_b + \lambda\eta_s \frac{z}{1+z}(v_0 - v_b - v_s)\Big] = \\ &= d + \gamma(v_s - v_0 + v_b) - \lambda\eta_s \frac{z}{1+z}(v_0 - v_b - v_s) \quad \Rightarrow \end{aligned}$$

$$\Rightarrow \quad v_0 - v_b = \frac{d + (\gamma + \lambda\eta_s \frac{z}{1+z})v_s}{r + \gamma + \lambda\eta_s \frac{z}{1+z}} \tag{A.20}$$

We can solve for v_s by plugging equation (A.20) into equation (A.19):

$$\begin{aligned} rv_s &= d - x + \lambda\eta_b \frac{1}{1+z}\left[\frac{d + \big(\gamma + \lambda\eta_s \frac{z}{1+z}\big)v_s}{r + \gamma + \lambda\eta_s \frac{z}{1+z}} - v_s\right] = \\ &= d - x + \lambda\eta_b \frac{d - rv_s}{(r + \gamma + \lambda\eta_s)z + r + \gamma} \quad \Rightarrow \\ &\Rightarrow \quad \left[1 + \frac{\lambda\eta_b}{(r + \gamma + \lambda\eta_s)z + r + \gamma}\right] rv_s = \left[1 + \frac{\lambda\eta_b}{(r + \gamma + \lambda\eta_s)z + r + \gamma}\right] d - x \quad \Rightarrow \end{aligned}$$

$$\Rightarrow \quad v_s = \frac{d}{r} - k\frac{x}{r} - k\frac{x}{(r + \gamma + \lambda\eta_s)z + \gamma} \tag{A.21}$$

where

$$k \equiv \frac{(r + \gamma + \lambda\eta_s)z + \gamma}{(r + \gamma + \lambda\eta_s)z + (r + \gamma + \lambda\eta_b)}$$

Given v_s, we can determined v_0, v_b and p uniquely from equations (8), (A.18) and (10) respectively. Let us compute them. We can solve for v_0 by plugging equation (A.21) into equation (8):

$$rv_0 = d + \gamma\left[\frac{d}{r} - k\frac{x}{r} - k\frac{x}{(r + \gamma + \lambda\eta_s)z + \gamma}\right] - \gamma v_0 \quad \Rightarrow$$

$$\Rightarrow \quad v_0 = \frac{d}{r} - k\frac{x}{r}\frac{\gamma}{r + \gamma} - k\frac{\gamma}{r + \gamma}\frac{x}{(r + \gamma + \lambda\eta_s)z + \gamma} \tag{A.22}$$

We now compute v_b by substituting equations (A.21) and (A.22) into equation (A.18):

$$
\begin{aligned}
rv_b &= -\gamma v_b + \lambda\eta_s \frac{z}{1+z}\Big[\frac{d}{r} - k\frac{x}{r}\frac{\gamma}{r+\gamma} - k\frac{\gamma}{r+\gamma}\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma} - \\
&- v_b - \frac{d}{r} + k\frac{x}{r} + k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big] \quad \Rightarrow \\
&\Rightarrow \quad \left[r+\gamma+\lambda\eta_s\frac{z}{1+z}\right]v_b = \lambda\eta_s\frac{z}{1+z}\left[k\frac{x}{r+\gamma} + k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\frac{r}{r+\gamma}\right] \quad \Rightarrow
\end{aligned}
$$

$$
\Rightarrow \quad v_b = k\frac{x}{r+\gamma}\frac{\lambda\eta_s z}{(r+\gamma+\lambda\eta_s)z+\gamma} \tag{A.23}
$$

We now solve for the price. Plugging equations (A.21) - (A.23) into equation (10) we get:

$$
\begin{aligned}
p &= \frac{1}{1+z}\Bigg[\Big(\frac{d}{r} - k\frac{x}{r} - k\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma}\Big)z + \\
&+ \frac{d}{r} - k\frac{x}{r}\frac{\gamma}{r+\gamma} - k\frac{\gamma}{r+\gamma}\frac{x}{(r+\gamma+\lambda\eta_s)z+\gamma} - k\frac{x}{r+\gamma}\frac{\lambda\eta_s z}{(r+\gamma+\lambda\eta_s)z+\gamma}\Bigg] = \\
&= \frac{d}{r} - \frac{1}{1+z}\left[k\frac{x}{r}\Big(\frac{\gamma}{r+\gamma} + z\Big) + k\frac{x}{r+\gamma}\right] \quad \Rightarrow
\end{aligned}
$$

$$
\Rightarrow \quad p = \frac{d}{r} - k\frac{x}{r} \tag{A.24}
$$

This concludes the proof of Proposition 4. □

B Additional Proofs

B.1 Proof of $\frac{\partial p}{\partial d}, \frac{\partial p}{\partial \eta_b} > 0$ and $\frac{\partial p}{\partial z}, \frac{\partial p}{\partial \eta_s} < 0$

Proof. Using equation (14), the partial derivative of the price with respect to the dividend flow d is

$$
\frac{\partial p}{\partial d} = \frac{1}{r} > 0 \quad \Rightarrow \quad \frac{\partial p}{\partial d} > 0 \quad \forall d
$$

Let us now compute the partial derivative of the price with respect to the measure of buyers-to-be η_b:

$$
\frac{\partial p}{\partial \eta_b} = \frac{\partial p}{\partial k}\frac{\partial k}{\partial \eta_b} = -\frac{x}{r}\frac{\partial k}{\partial \eta_b}
$$

where:

$$\frac{\partial k}{\partial \eta_b} = -\lambda \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}$$

which is strictly lower than zero since $r, \gamma, \lambda, \eta_s, z > 0$. Therefore,

$$\frac{\partial p}{\partial \eta_b} > 0 \quad \forall \eta_b$$

Next, we obtain the partial derivative of the price with respect to the buyer's-to-be bargaining power z:

$$\frac{\partial p}{\partial z} = \frac{\partial p}{\partial k}\frac{\partial k}{\partial z} = -\frac{x}{r}\frac{\partial k}{\partial z}$$

where:

$$\frac{\partial k}{\partial z} = \frac{r\lambda\eta_s\eta_b}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}$$

which is strictly greater than zero since $r, \lambda, \eta_s, \eta_b > 0$. Then,

$$\frac{\partial p}{\partial z} < 0 \quad \forall z$$

To complete the proof, we calculate the partial derivative of the asset price with respect to the measure of sellers-to-be:

$$\frac{\partial p}{\partial \eta_s} = \frac{\partial p}{\partial k}\frac{\partial k}{\partial \eta_s} = -\frac{x}{r}\frac{\partial k}{\partial \eta_s}$$

where:

$$\frac{\partial k}{\partial \eta_s} = \frac{\lambda z(r+\lambda\eta_b)}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}$$

which is strictly greater than zero since $r, \lambda, \eta_b, z > 0$. Thus,

$$\frac{\partial p}{\partial \eta_s} < 0 \quad \forall \eta_s$$

□

B.2 Proof of $\frac{\partial p}{\partial g} > 0$

Proof. Using equation (14), the partial derivative of the asset price with respect to the flow of investors g entering the market is

$$\frac{\partial p}{\partial g} = -\frac{x}{r}\frac{\partial k}{\partial g}$$

where $k = \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$. Let us derive the partial derivative of k with respect to the flow of investors g:

$$\frac{\partial k}{\partial g} = \frac{1}{[(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)]^2}\left\{\Big(r+\lambda\eta_b\Big)\lambda z\frac{\partial \eta_s}{\partial g} - \Big[\Big(r+\gamma+\lambda\eta_s\Big)z+\gamma\Big]\lambda\frac{\partial \eta_b}{\partial g}\right\}$$

which is strictly lower than zero since $r, \gamma, \lambda, \eta_s, \eta_b, z > 0$ and, as shown in Proposition 2, $\frac{\partial \eta_s}{\partial g} < 0$ and $\frac{\partial \eta_b}{\partial g} > 0$. Then,

$$\frac{\partial p}{\partial g} = -\frac{x}{r}\frac{\partial k}{\partial g} > 0$$

which proves the price increases in the flow of investors entering the market. □

C Proof of Theorem 1

Proof. In our framework, the marginal investor decides whether to enter or not after comparing the expected utility v_{alt} of investing in an alternative market to the expected utility v_b of a buyer-to-be in our market. The expected utility of the marginal investor $v_{alt} = \kappa'$ is a non-negative and strictly increasing function of his outside investment opportunity κ'. Also, $v_b(\kappa' = 0) > v_{alt}(\kappa' = 0) = 0$. Hence, if v_b were decreasing in the outside investment opportunity of the marginal investor, κ', then there would be a unique threshold κ^* satisfying the indifference condition $v_b(g(\kappa^*)) = v_{alt}(\kappa^*)$. Let us show this is the case.

The expected utility v_b of a buyer-to-be is a function of the flow of investors g entering the market. Let us compute the partial derivative of v_b, defined in equation (11), with respect to g:

$$\frac{\partial v_b}{\partial g} = \frac{\lambda z x}{r+\gamma}\frac{1}{\left[(1+z)(r+\gamma)+\lambda(z\eta_s+\eta_b)\right]^2}\left\{\Big[(1+z)(r+\gamma)+\lambda\eta_b\Big]\frac{\partial \eta_s}{\partial g} - \lambda\eta_s\frac{\partial \eta_b}{\partial g}\right\}$$

which is strictly negative since $r, \gamma, x, z, \lambda, \eta_b, \eta_s > 0$ and, as shown in Proposition 2, $\frac{\partial \eta_s}{\partial g} < 0$ and $\frac{\partial \eta_b}{\partial g} > 0$. Hence, the expected utility v_b of a buyer-to-be strictly decreases in the flow of investors g entering the market. However g, as given by $g(\kappa') = \int_{\underline{\kappa}}^{\overline{\kappa}} \nu(\kappa) f(\kappa) d\kappa = \int_{\underline{\kappa}}^{\kappa'} f(\kappa) d\kappa$, is increasing in κ'. As a result,

$$\frac{\partial v_b}{\partial \kappa'} = \frac{\partial v_b}{\partial g}\frac{\partial g}{\partial \kappa'} \leq 0$$

where $\frac{\partial v_b}{\partial g} < 0$ and $\frac{\partial g}{\partial \kappa'} \geq 0$.

Then, by continuity, there exists a unique value of κ' satisfying the indifference condition: $v_b(g(\kappa^*)) = v_{alt}(\kappa^*)$. A unique threshold κ^* thus defines a unique flow of investors $g^* = g(\kappa^*)$ entering the market. But given a flow of investors entering the market, there exists unique equilibrium measures (η_b^*, η_0^*, η_s^*) of each type of investor, expected utilities (v_b^*, v_0^*, v_s^*) and price of the asset, p^*, as proved in Propositions 1 and 4. Consequently, market equilibrium, as presented in Definition 1, is unique. This proves Theorem 1. □

D Proof of Proposition 6

Proof. The illiquidity discount is defined as:

$$k\frac{x}{r}$$

where $\frac{x}{r}$ is the present value of the holding cost x, $k = \frac{(r+\gamma+\lambda\eta_s)z+\gamma}{(r+\gamma+\lambda\eta_s)z+(r+\gamma+\lambda\eta_b)}$ and η_s and η_b, as given by equations (5) and (6), are functions of the flow of investors g. Let us compute the partial derivative with respect to the flow of investors g entering the market:

$$\frac{\partial}{\partial g}\left(k\frac{x}{r}\right) = \frac{x}{r}\frac{\partial k}{\partial g} < 0$$

since $\frac{\partial k}{\partial g} < 0$ (as shown in subsection B.2) and $x, r > 0$. Hence, illiquidity decreases in g or equivalently, market liquidity increases in the flow of investors entering our market. □

References

Acharya, V. V. and L. H. Pedersen (2005). Asset Pricing with Liquidity. *Journal of Financial Economics 77*(2), 375–410.

Amihud, Y. and H. Mendelson (1986). Asset Pricing and Bid-ask Spread. *Journal of Financial Economics 17*(2), 223–249.

Amihud, Y., H. Mendelson, and L. H. Pedersen (2005). Liquidity and Asset Prices. *Foundations and Trends in Finance 1*(4), 269–364.

BIS (1999, October). A Review of Financial Market Events in Autumn 1998. *CGFS Publication* (12). http://www.bis.org/ publ/ cgfs12.htm.

Brunnermeier, M. K. and L. H. Pedersen (2007). Market Liquidity and Funding Liquidity. *Mimeo, Princeton University*. http://www.princeton.edu/ ∼markus/ research/ papers/ liquidity.

Burden, R. L. and J. D. Faires (1993). *Numerical Analysis*. Boston (Massachusetts): PWS Publishing Company.

Cohen, B. H. and E. M. Remolona (2001, December). Overview: Financial Markets Prove Resilient. *BIS Quarterly Review*. http://www.bis.org/ publ/ r_qt0112a.pdf.

Cooper, R. and A. John (1988). Coordinating Coordination Failures in Keynesian Models. *The Quarterly Journal of Economics 103*(3), 441–463.

Cooper, R. W. (1999). *Coordination Games: Complementarities and Macroeconomics*. Cambridge (United Kingdom): Cambridge University Press.

Diamond, P. A. (1982a). Aggregate Demand Management in Search Equilibrium. *Journal of Political Economy 90*(5), 881–894.

Diamond, P. A. (1982b). Wage Determination and Efficiency in Search Equilibrium. *The Review of Economic Studies 49*(2), 217–227.

Dow, J. (2004). Is Liquidity Self-Fulfilling? *Journal of Business 77*(4), 895–908.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2005). Over-the-Counter Markets. *Econometrica 73*(6), 1815–1847.

Duffie, D., N. Gârleanu, and L. H. Pedersen (2007). Valuation in Over-the-Counter Markets. *Review of Financial Studies*, forthcoming.

Duffie, D. and Y. Sun (2007). Existence of Independent Random Matching. *Annals of Applied Probability 17*(1), 386–419.

Evans, M., N. Hastings, and B. Peacock (1993). *Statistical Distributions (Second edition).* New York: Wiley-Interscience.

Fender, I. and P. Hördahl (2007). Overview: Credit Retrenchment Triggers Liquidity Squeeze. *BIS Quarterly Review*. http://www.bis.org/ publ/ qtrpdf/ r_qt0709a.pdf.

Gromb, D. and D. Vayanos (2002). Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs. *Journal of Financial Economics 66*(2-3), 361–407.

Huang, J. C. and J. Wang (2007). Market Liquidity, Asset Prices, and Welfare. *Mimeo, McCombs School of Business*. http://papers.ssrn.com/ sol3/ papers.cfm?abstract_id=1009568.

IMF (1998, December). World Economic Outlook and International Capital Markets: Interim Assessment. *Mimeo, McCombs School of Business*. http://www.imf.org/ external/ pubs/ ft/ weo/ weo1298/ index.htm.

Lagos, R. and G. Rocheteau (2007). Liquidity in Asset Markets with Search Frictions. *Mimeo, New York University*. http://rl561a.googlepages.com/ workingpapers.

McAndrews, J. J. and S. M. Potter (2002). Liquidity Effects of the Events of September 11, 2001. *Federal Reserve Bank of New York Economic Policy Review*, 59–79.

Milgrom, P. and J. Roberts (1990). Rationalizability, Learning and Equilibrium in Games with Strategic Complementarities. *Econometrica 58*(6), 1255–77.

Mortensen, D. T. (1982, J.J. Mccall, ed.). The Matching Process as a Non-Cooperative/Bargaining Game. *The Economics of Information and Uncertainty New York: NBER Conference Volume.*

Pagano, M. (1989). Trading Volume and Asset Liquidity. *The Quarterly Journal of Economics 104*(2), 255–274.

Phelps, E. (1972). *Microeconomic Foundations of Employment and Inflation Theory.* New York: W. W. Norton & Company, Inc.

Pissarides, C. A. (1985). Short-Run Equilibrium Dynamics of Unemployment, Vacancies, And Real Wages. *American Economic Review 75*(4), 676–690.

Pissarides, C. A. (2001). Economics of Search. *International Encyclopedia of the Social & Behavioral Sciences*, 13760–13768.

Plantin, G. (2004). Self-Fulfilling Liquidity and the Coordination Premium. *Mimeo, London Business School*. http://ideas.repec.org/p/cmu/ gsiawp/ 710428070.html.

Topkis, D. M. (1979, November). Equilibrium Points In Nonzero-Sum n-Person Submodular Games. *SIAM Journal on Control and Optimization 17*(6), 773–787.

Vayanos, D. (1998). Transaction Costs and Asset Prices: A Dynamic Equilibrium Model. *Review of Financial Studies 11*(1), 1–58.

Vayanos, D. and T. Wang (2007). Search and Endogenous Concentration of Liquidity in Asset Markets. *Journal of Economic Theory*, forthcoming.

Vayanos, D. and P.-O. Weill (2007). A Search-Based Theory of the On-the-Run Phenomenon. *Journal of Finance*, forthcoming.

Weill, P.-O. (2007). Liquidity Premia in Dynamic Bargaining Markets. *Mimeo, UCLA*. http://www.econ.ucla.edu/ poweill/ liquiditypremia.pdf.

Liquidity and Transparency in Bank Risk Management

Lev Ratnovski *

University of Amsterdam and Bank of England

February 2008

Abstract

Banks are exposed to liquidity risk when solvency concerns arise at the refinancing stage. To manage that risk, banks can accumulate liquid assets, or enhance transparency to facilitate refinancing. A liquidity buffer provides complete insurance against small liquidity needs, while transparency offers partial insurance against large ones as well. Without regulatory incentives, banks can under-invest in both liquidity and transparency. While liquidity can be imposed, transparency is not verifiable. This creates a multi-tasking problem in liquidity regulation. Liquidity requirements can compromise banks' endogenous transparency choices, leaving them exposed to large shocks. Liquidity risks can increase in response to seemingly more stringent regulation.

*I thank for useful comments Viral Acharya, Arnoud Boot, Charles Goodhart, Mark Flannery, Charles Kahn, Erlend Nier, Per Ostberg, Daniel Paravisini, Enrico Perotti, Rafael Repullo, Jean-Charles Rochet, Javier Suarez, Ernst-Ludwig von Thadden, Tanju Yorulmazer, and participants of the WFA meetings (2006, Denver), EFA meetings (2006, Zurich), Basel Committee workshop on "Banking, Risk, and Regulation" at FDIC, CEPR conference on "Corporate Finance and Risk Management", JFS/Bank of Finland conference "Financial Instability, Supervision and Central Banks", and LSE conference on "Cycles, Contagion, and Crises". All errors are mine. The views expressed are those of the authors and do not necessarily reflect those of the Bank of England. Contact: ratnovski@gmail.com.

1 Introduction

Banks perform maturity transformation: they use short-term liabilities (demandable and term) to finance predominantly long-term assets. They therefore need to routinely roll over maturing debt, and refinance or cover from precautionary liquidity buffers any withdrawals that do occur. A bank unable to do so may fail in a liquidity shortage despite being long-term solvent[1]. This is the origin of bank liquidity risk.

The majority of recent liquidity events in developed economies were caused by a sudden increase in uncertainty over banks' solvency, which prevented normally smooth intermediate refinancing. This applied both to idiosyncratic crises, when a specific bank could not refinance despite relatively abundant market liquidity, and to systemic liquidity events:

- *Citibank* and *Standard Chartered* (HK) in 1991: rumors of technical insolvency;
- *Lehman Brothers* in 1998: rumors of severe losses in emerging markets;
- *Commerzbank* in 2002: rumors of insolvency triggered by trading losses;
- The credit and interbank market turmoil of 2007– is admittedly complex in nature. Yet it is also in large part attributed to uncertainty over the solvency of banks involved (due to hard-to-value subprime and off-balance sheet exposures).

Banks affected by liquidity problems typically had to cover most maturing obligations with minimal access to new funds. The strain was most severe in wholesale funding. Retail outflows, on the contrary, were relatively modest even during most publicized events: only 5% of deposit base for BAWAG in 2006 and 8% for Northern Rock in 2007.

The purpose of this paper is to offer a model of liquidity risk consistent with these stylized facts, and to study consequent risk management and regulatory implications. Our principal observation is that a bank can insure liquidity risk caused by uncertainty

[1] Here and further we use the term "solvency" to describe a bank with valuable long-term assets and hence positive capital.

over its solvency in two ways. One, traditional, is to accumulate a precautionary buffer of liquid assets. Another, less conventional, is to create conditions that facilitate its access to external refinancing. We associate the latter with investment in transparency, and analyze the interaction between bank's liquidity buffers and transparency choices.

In the model, we consider a bank with a valuable long-term project that normally produces a high return, but with a small probability can turn out to be of zero value. Because solvency risk is small, that does not prevent initial funding. At the intermediate date, the bank needs to refinance an exogenous random withdrawal. Its ability to do so can be compromised by informational frictions. In most states of the world, the bank is confirmed to be solvent, and investors are willing to refinance it. Yet with some probability investors receive a noisy negative signal that the likelihood of insolvency is high. In that case, investors will become unwilling to refinance, creating liquidity risk – and a possibility of failure, even for a solvent bank. The effect where the lemon premium can suddenly increase during the refinancing stage is the key feature of this model. It allows explaining why, with no change in fundamental value, banks that could easily obtain initial funding may become exposed to refinancing frictions later.

When a bank experiences intermediate withdrawals, a precautionary *liquidity buffer* of easily tradeable short-term assets allows it to cover possible withdrawals internally. An alternative hedging strategy is for the bank to adopt *transparency*. We understand transparency as a set of mechanisms that make the value of banks' assets more observable at the intermediate stage. That reduces the probability of refinancing frictions in solvent banks, allowing them to borrow from the market to substitute withdrawals. Both liquidity and transparency are ex-ante investments that need to be undertaken before intermediate shocks realize.

While liquidity and transparency are strategic substitutes, their precise effects are different. A precautionary liquidity buffer allows the bank to cover any outflows within its size, providing complete insurance against small liquidity needs. Transparency, on the other hand, helps resolve solvency concerns and obtain external refinancing for liquidity needs of any size. Yet it is effective only with probability (ex-post communication is imperfect) and hence provides incomplete insurance. This leads to the result that liquidity and transparency can complement each other. Banks can combine them in risk management, using liquidity buffers to fully insure against small shocks, and transparency (enhanced ability to borrow) to partially cover large shocks as well.

Banks' incentives to invest in liquidity and transparency can be distorted by leverage (Jensen and Meckling, 1976). Suboptimal hedging justifies policy intervention. However, while liquidity buffers are verifiable and can be imposed (for example, through explicit ratios), transparency is not easily verifiable and is more difficult to regulate. We show that this leads to a multi-tasking problem, where imposed liquidity requirements can compromise banks' transparency choices. The reason is that more liquid banks are insured against a wider range of shocks and have lower marginal benefits of investing in transparency. When transparency deteriorates, banks become more exposed to large liquidity needs. As a result, under certain conditions, making banks more liquid may actually *increase* liquidity risks they face.

Our analysis yields a number of topical policy implications. Firstly, we caution on the consequences of wrongly designed liquidity requirements. They can have unwanted effects, such as reduced banks' transparency. In the extreme, this can lead to a net increase in liquidity risk as a result of seemingly more stringent regulation. Traditionally, liquidity regulation centers around assuring necessary liquidity buffers. Our analysis suggests that the considerations of transparency and market access should be another prominent dimension.

Secondly, the model demonstrates that solvency regulation alone cannot fully address liquidity risk. The reason is that refinancing frictions are driven by *imperfect information* on solvency. Unless more precise information is available, higher solvency would not by itself necessarily lead to a reduction in liquidity risk. And even if it does, using liquidity regulation (rather than excessively stringent solvency requirements) can be more efficient.

Finally, the model sheds light on the relationship between the move to market-based funding of banks (and the proliferation of non-bank financial intermediaries) and liquidity risks in the financial system. If liquidity risk was primarily driven by demandable deposits, then lower reliance on retail funding would imply reduced risks. Yet, our model shows how liquidity risk can arise without demandable deposits and classic bank runs. In fact, extrapolating from recent events, retail liabilities are a relatively stable source of bank funding (cf. Gatev and Strahan, 2006). Consequently, liquidity risks are likely to remain, if not increase, as financial intermediaries move to market-based funding. Indeed, during the turmoil of 2007–, banks that most heavily relied on wholesale funding (such as Countrywide in the US or Northern Rock in the UK) appeared particularly vulnerable.

This paper contributes to the literature on bank liquidity risk. It is most closely related to models of liquidity risk driven by asymmetric information. In a seminal paper, Chari and Jagannathan (1987) consider consumer-based panics. When uninformed depositors make solvency inferences by observing withdrawals of others, but cannot distinguish between information- and liquidity-based ones, they can run on fundamentally solvent banks. Interestingly, Chari and Jagannathan explicitly rule out informed refinancing. They make clear that "the most serious problem" with their approach is that it assumes "the absence of markets for trading is asset claims" (*p.722, remark 2*).

Goodfriend and King (1988) argue that, under efficient inter-bank markets, a bank known to be solvent should always be able to refinance itself. This provides a useful benchmark. A number of consequent papers demonstrate how information imperfections can prevent refinancing of solvent banks. Flannery (1996) argues that if potential lenders are uncertain of their screening ability, they can all withdraw from the market to avoid being exposed to prohibitive lemon costs. Rochet and Vives (2004) show a possibility of a coordination failure among potential lenders, where each of them does not lend if expects the same from others. The paper probably closest to ours is Freixas et al. (2004), who model wholesale market frictions under asymmetric information on solvency. They assume that a solvent bank that requires refinancing is indistinguishable from an insolvent one that attracts additional funds to gamble for resurrection.

In this context, our paper has two principal novelties. The first is to offer a streamlined approach to modelling liquidity risks, where both solvent and insolvent banks face a fundamentally identical liquidity need (to refinance withdrawals), while refinancing frictions are driven by lemon costs imposed on solvent banks that are by chance indistinguishable from insolvent ones. Secondly, we have a specific focus on the analysis of risk management, formalizing the interaction between liquidity buffers and transparency (ability to borrow) and examining implications for ex-ante regulation.

The paper proceeds as follows. Section 2 discusses the notion of bank transparency. Section 3 sets up the model of liquidity risk. Section 4 introduces hedging choices, and studies the interaction between liquidity buffers and transparency. Section 5 observes that banks can make suboptimal hedging choices. Section 6 examines multi-tasking in liquidity regulation. Section 7 discusses robustness and empirical implications. Section 8 concludes.

2 What is Bank Transparency?

The notion of bank transparency has received significant attention in policy debate. It is closely linked to Pillar 3 (market discipline) of Basel II. Most recently, boosting banks' transparency was seen as a crucial condition for resolving and preventing the recurrence of the 2007– credit turmoil[2]. How to think about bank transparency?

Banks are inherently opaque, because through lending relationships they obtain non-verifiable information on borrowers. Banks can use ex-post disclosure (information release) to communicate with outsiders, but without appropriate preconditions mechanical disclosure can be not credible or ineffective (Boot and Thakor, 2001). Transparency can

[2] "*A roadmap of reforms to boost transparency ... is being drawn up by eurozone finance ministers to prevent a repeat of turmoil*", FT, Oct 09, 2007; "*Mr Trichet said other lessons to be drawn included the need to increase financial market transparency*", FT, Nov 23, 2007; "*At a meeting in Paris yesterday to discuss their response to the credit squeeze, the ministers from France, Germany, Italy and the UK agreed to step up efforts to improve transparency*", FT, Jan 18, 2008; "*Jean-Claude Trichet, president of the European Central Bank, admitted: 'Transparency is going to be crucial'*", FT, Jan 25, 2008.

be seen as a set of ex-ante choices that facilitate effective disclosure ex-post (Perotti and Von Thadden, 2003).

There are two ways by which a bank can improve information available to outsiders. One is to offer less information-sensitive loans (Boot and Thakor, 2000), for example standardized mortgages. The cost of such choice is that less information-sensitive loans may have lower margins, for example due to higher competition and lower relationship rents. Another possibility, which is the focus of this paper, is to make specific (and costly) investments *in transparency* that make asset quality more observable for any given asset mix. For example:

- Issue subordinated debt (Calomiris, 1999) even when cheaper funding is possible, in order to have market participants who specialize on assessing the bank available should a genuine need for funds arise;
- Invest in risk management and accounting systems that produce better (and externally verifiable) information on the effects of possible shocks;
- Streamline "large and complex financial institutions" to make their individual businesses more comparable with specialized counterparts;
- Build reputation for credible disclosure[3].

There is no universal "recipe" for achieving transparency. Neither of the above mechanisms is sufficient by itself. For example, mandatory disclosure is ineffective when it is difficult to define relevant quantifiable parameters, or when banks can engage in "creative" reporting. While the subordinated debt suggestion of Calomiris (1999) is intriguing, it has not yet been fully tested in practice[4]. This makes transparency hard

[3] For an interesting illustration of this see Griffin and Wallach (1991). In late 1980-es, many U.S. banks experienced large and uncertain losses on defaulted Latin American debts. That hampered their access to credit markets. In May 2007, Citicorp became the first large bank to make substantial provisions ($3 billion). That served as a signal of commitment to draw a line under prior losses – credible due to the risk of a negative market reaction if realized losses ended up being higher. The provisioning led to a positive market reaction and improved access to funds.

[4] The ABCP crisis of 2007 demonstrated how debt valuations can deviate from fundamentals. Although ABCP was senior, it is not evident that junior debt is immune against similar distortions.

to contract upon or to regulate, especially when the measures required to improve it differ across banks.

The last remark concerns the relationship between asset liquidity and transparency. Liquid and actively traded assets can be regarded as transparent when their price is publicly known and provides relatively precise information on fundamental value. During periods of mispricing, assets may be liquid but not transparent. Also, not all transparent assets are liquid. For example, even when outsiders have good information on assets' fundamental value, but collecting repayments requires bank-specific skills (cf. Diamond and Rajan, 2001), such assets cannot be simply sold. Instead, a bank may raise cash by borrowing against their future value. This is the dichotomy between holding liquid assets that can be simply converted into cash and having transparency over the value of illiquid assets that enables borrowing, which we explore in this paper.

3 A Model of Liquidity Risk

3.1 Economy and Agents

Consider a risk-neutral economy with three dates $(0, 1, 2)$ and no discounting. The economy is populated by multiple competitive investors and a single bank. The investors are endowed with money that they can lend to the bank against expected rate of return 1.

The bank has no initial capital, but enjoys exclusive access to a profitable investment project. For each unit of financing at date 0, the project returns at date 2 a high return X with probability $1 - s$, but 0 with a small probability s (s for the probability of a solvency problem). The bank operates under a leverage constraint and cannot borrow more than 1 at date 0. It is financed by debt (some of it is short-term and needs to be refinanced at date 1, as detailed below) and maximizes date 2 profit. The timeline is given in Figure 1.

<< *Figure 1* >>

3.2 Solvency Concerns and Liquidity Risk

Two events happen at date 1. One is a random withdrawal of a part of initial funding. Another is a signal on bank solvency. The two events are independent – withdrawals are made by uninformed depositors or represent maturing term funding, and are therefore not influenced by the solvency signal.

Withdrawals and liquidity need While the project is long-term, some debt matures earlier and must be refinanced. In reality there may be multiple refinancing events through the course of the project, but for the analysis we collapse them into a single "intermediate" date 1. The amount of funds maturing at date 1 – liquidity need – is random. With probability l, the liquidity need is low: the bank has to repay some $L < 1$. With additional probability $1 - l$, the liquidity need is high: the bank has to repay 1. If a bank cannot repay, it fails and goes bankrupt with no liquidation value.

Information and liquidity risk Because investors always offer an elastic supply of funds (there is no aggregate liquidity shortage in this model), a bank known to be solvent can refinance any withdrawals by new borrowing. Yet smooth refinancing can be impeded by the effects of imperfect information, namely – increased solvency concerns. That is the origin of liquidity risk, and the key ingredient of this model.

Recall that a bank is fundamentally solvent with probability $1-s$ and insolvent with probability s. Assume that, at date 1, investors receive a *noisy* signal of bank solvency. Concretely, with probability $1-(s+q)$ there is a correct "positive" signal that a bank is solvent and will yield X with certainty. Solvent banks are able to refinance themselves at a risk-free rate.

However, with a residual probability $s+q$, there is a "negative" signal that a bank is likely to be insolvent. That signal is received by a mass s of genuinely insolvent banks, but also by a mass q of solvent banks that are wrongly pooled together with insolvent ones. The posterior probability of insolvency under a "negative" signal, $s/(s+q)$, is higher then the ex-ante probability of insolvency, s. Higher solvency risk can prevent external refinancing, creating liquidity risk. The informational structure determining liquidity risk in this model is illustrated in Figure 2.

<< *Figure 2* >>

The model allows different interpretations of the availability of solvency information to bank managers. The only constraint is that all banks, independently of fundamental solvency, should have incentives to seek refinancing. The simplest approach is to assume that bank managers obtain the same signal as outsiders. (The fact that insiders themselves lack information on solvency is in line with the observations of the 2007– turmoil, when some banks had to revise loss estimates or uncovered previously unexpected losses.) Alternatively, bank managers could receive a more precise but still imperfect signal (so that a small probability of success remains even upon receiving a negative signal), or simply enjoy private benefits of running a bank. In either case, all banks will choose to seek refinancing since any probability of obtaining it dominates the zero payoff in case of immediate liquidation at date 1.

The fact that the lemon premium for a bank affected by solvency concerns at date 1 is higher than the original lemon premium at date 0 is the key to our modelling of liquidity risk. A sudden increase in lemon premium can prevent refinancing even though a bank was able to attract initial funding.

Formally, we impose two restrictions. Firstly:

$$X > \frac{1}{1-(s+q)} \tag{A1}$$

This assures that a bank can always obtain initial financing at date 0. Even if it always failed in a liquidity shock (upon a "negative" signal at date 1, which happens with probability $1-(s+q)$), it could borrow by offering repayment $1/[1-(s+q)]$ in case of success. Secondly:

$$X < (1-L) + L \cdot \frac{s+q}{q} \tag{A2}$$

This is a sufficient condition under which a bank with a "negative" signal at date 1 cannot obtain refinancing. The condition addresses the most mild scenario of a small withdrawal of size L, to refinance which the bank has to offer repayment $L \cdot [s+q]/q$, and the lowest possible interest rate 1 on the rest $(1-L)$ of initial funding which does not need to be refinanced. A bank faced with a more severe scenario of a large withdrawal of size 1 would also be unable to refinance: from (A2), $X < [s+q]/q$.

Observe that there exist parameter values such that (A1) and (A2) are satisfied simultaneously (take $s+q << 1$ and $q << s$). We can now summarize the main property of this model:

Proposition 1 *Under (A1) and (A2), a bank can attract initial funding at date* 0*, but if faced with solvency concerns (in a mass q of solvent banks pooled together with a mass s of insolvent banks) cannot obtain intermediate refinancing at date* 1*. This is the source of liquidity risk in the model.*

The corollary from Proposition 1 is that the ability to separate from insolvent banks (which we will further interpret as transparency) is less important at date 0 but may become critical in case of a "negative" intermediate signal at date 1.

Before proceeding to the analysis of risk management options, we make a simplifying assumption that initial financing at date 0 is covered by deposit insurance. This means that the bank has to promise original investors the repayment of 1. To preserve refinancing frictions at the intermediate stage, we maintain that date 1 refinancing is *not* covered by deposit insurance. (This is plausible e.g. when date 0 investments are deposits, while date 1 refinancing is market-based, corresponding to the practice of using wholesale funds to manage liquidity needs.) The deposit insurance assumption does not affect qualitative properties of the model; it reduces leverage (which is the main distortion) and can only weaken our results. We discuss robustness in Section 7.

4 Liquidity Risk Management

In this section, we introduce the instruments of bank liquidity risk management – liquidity buffers and transparency – and analyze socially optimal hedging choices. We show that liquidity and transparency are only partial substitutes, and for low enough costs of hedging can be optimally combined in risk management. Then, a precautionary liquidity buffer fully insures a solvent bank against small withdrawals, which happen with probability l. At the same time, transparency partially insures against large withdrawals as well, by allowing to confirm bank solvency and enabling external refinancing with probability t.

4.1 Instruments

We consider two ways in which a bank can hedge its liquidity risk.

Firstly, a bank can *accumulate a liquidity buffer.* A bank can invest L into short-term assets (storage: cash or easily tradeable securities that at any date produce a safe but minimal return of 1). This allows to fully cover possible small withdrawals at date 1, which happen with probability l. Note that, by construction, a bank cannot use liquidity to insure against large withdrawals (of size 1) because that would require allocating all initial financing to storage, leaving nothing for the profitable investment.

Secondly, a bank can *adopt transparency.* It needs to spend T to establish it. We think of transparency as a strategic ex-ante investment that facilitates future information communication. Transparency can help the bank publicly confirm its solvency and refinance both small and large liquidity needs. Yet, transparency is imperfect due to unavoidable frictions in ex-post communication; we assume that it is effective only with probability $t < 1$.[5]

Both liquidity and transparency have costs. They crowd out profitable investment.

[5] The effectiveness of bank's investment in transparency can be affected by country-specific (more developed and liquid financial markets) or industry-specific (transparent peer banks or location in the financial centre when there are positive informational externalities, Admati and Pfleiderer, 2000) factors that are outside a single bank's control.

Investing L in liquidity reduces return to a successful bank by $L(X-1)$, investing T in transparency – by TX, and having both liquidity and transparency together – by $L(X-1)+TX$. Since we are primarily interested in different hedging effects of liquidity and transparency, we take their costs to be equal: $L(X-1) = TX = C$ (for both hedges $L(X-1) + TX = 2C$), where C is the generic cost of hedging. We also assume that, for simplicity, when a liquid bank fails at date 1 its liquidity buffer is lost (liquidity is allocated to costly bankruptcy proceedings or converted into marginal bankers' private benefits, Myers and Rajan, 1998); this makes return to a failing bank always 0.

For definitiveness, when indifferent, banks prefer to be hedged, and prefer liquidity over transparency. Bank returns, depending on its ex-ante hedging choice and the shock realized at date 1, are summarized in Figure 3.

<< *Figure 3* >>

4.2 Hedging Strategies

We first derive the levels of social welfare corresponding to different bank hedging choices.

When a bank is neither liquid nor transparent (strategy "N"):

$$\Pi_N^S = (1 - s - q) \cdot X - 1$$

Here, $1 - s - q$ is the probability that a bank is not hit by a solvency or liquidity shock, X is the return in that case, and 1 is the initial investment.

When a bank is liquid but not transparent (strategy "L"):

$$\Pi_L^S = (1 - s - q(1 - l)) \cdot (X - C) - 1$$

A solvent bank survives a small liquidity shock (probability ql) by covering it from the precautionary buffer. It fails in a solvency shock (probability s) or in a large liquidity shock when withdrawals exceed the size of the buffer (probability $q(1 - l)$). Therefore the probability of survival is $1 - s - q(1 - l)$, the return in that case is $X - C$ (C is the cost of hedging), and the initial investment is 1.

When a bank is transparent but not liquid (strategy "T"):

$$\Pi_T^S = (1 - s - q(1 - t)) \cdot (X - C) - 1$$

A solvent bank survives a liquidity shock (either small or large) when it is successful in communicating solvency information to the market, with probability t. It fails in a solvency shock (probability s) or in a liquidity shock when transparency is ineffective (probability $q(1-t)$). Therefore the probability of survival is $1-s-q(1-t)$, the return in that case is $X - C$, and the initial investment is 1.

Lastly, when a bank is both liquid and transparent (strategy "LT"):

$$\Pi_{LT}^S = (1 - s - q(1 - l)(1 - t)) \cdot (X - 2C) - 1$$

A solvent bank survives a small liquidity shock (probability ql) by covering it from a precautionary buffer, and a large liquidity shock when successful in communicating solvency information, with probability t. It fails in a solvency shock (probability s) or in a large liquidity shock when transparency is ineffective (probability $q(1 - l)(1 - t)$). Therefore, the probability of survival is $1 - s - q(1 - l)(1 - t)$, the return in that case is $X - 2C$ (note double hedging cost), and the initial investment is 1.

4.3 Optimal Risk Management

We use these four payoffs to compare social welfare and derive bank's socially optimal hedging strategy.

Consider first the choice between liquidity and transparency. Liquidity insures a share l of shocks – small ones only. Transparency insures a share t of shocks – when ex-post information communication is successful. Thus for $l \geq t$ liquidity is more effective: $\Pi_L^S \geq \Pi_T^S$, and for $l < t$ transparency is more effective: $\Pi_L^S < \Pi_T^S$.

Another dimension is the depth of hedging – whether to hedge at all, adopt a single hedge, or have both hedges. Note that the marginal benefit of having a second hedge is lower than that of the first hedge. This is because the first hedge adopted is a more effective one (liquidity for $l \geq t$ and transparency for $l < t$), and already protects a bank from a range of liquidity shocks. Optimal depth of hedging depends on its cost C. It is optimal that a bank has no hedge for high costs of hedging, a single hedge (liquidity or transparency, whichever more effective) for intermediate costs of hedging, and both hedges (liquidity *and* transparency) for low costs of hedging. We consider two cases:

Case 1: Liquidity is more effective, $l \geq t$. It is optimal that a bank:

– Has no hedge, "N", for $\Pi_N^S > \Pi_L^S$, corresponding to high costs of hedging:

$$C > \frac{ql}{1 - s - q(1-l)} \cdot X$$

– Is only liquid, "L", for $\Pi_L^S \geq \Pi_N^S$ and $\Pi_L^S > \Pi_{LT}^S$, corresponding to intermediate costs of hedging:

$$\frac{q(1-l)t}{1 - s - q(1-l)(1-2t)} \cdot X < C \leq \frac{ql}{1 - s - q(1-l)} \cdot X$$

– Is both liquid and transparent, "LT", for $\Pi_{LT}^S \geq \Pi_L^S$, corresponding to low costs of hedging:

$$C \leq \frac{q(1-l)t}{1 - s - q(1-l)(1-2t)} \cdot X \tag{1}$$

Case 2: Transparency is more effective, $l < t$. Analogously, it is optimal that a bank:

– Has no hedge, "N", for $\Pi_N^S > \Pi_T^S$:

$$C > \frac{qt}{1 - s - q(1-t)} \cdot X$$

– Is only transparent, "T", for $\Pi_T^S \geq \Pi_N^S$ and $\Pi_T^S > \Pi_{LT}^S$:

$$\frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot X < C \leq \frac{qt}{1-s-q(1-t)} \cdot X$$

– Is both liquid and transparent, "LT", for $\Pi_{LT}^S \geq \Pi_T^S$:

$$C \leq \frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot X \tag{2}$$

The menu of equilibrium hedging outcomes (for a given value of C, depending on l and t) is illustrated in Figure 4. Observe that a bank chooses to remain uninsured when the effectiveness of hedging parameters l and t are relatively low. A bank chooses a single hedge for low-to-medium values of l and t, and both hedges for medium-to-high values of l and t. Note that for very high l or t a bank is better off remaining with a single most effective hedge.

<< *Figure 4* >>

Now consider conditions (1) and (2). Observe that they have strictly positive right sides. Therefore, for any l, t, q, s, there exists a cost of hedging C low enough, such that any of them holds, and it is socially optimal for a bank to be both liquid and transparent. This leads to the following main result:

Proposition 2 *When costs of hedging are low enough, it is optimal that banks combine liquidity and transparency in their risk management. For any l, t, q, s, there exists C low enough, such that conditions (1) for $l \geq t$ or (2) for $l < t$ are satisfied.*

Proposition 2 establishes that there exist conditions when it is optimal for the bank to be both liquid (to fully hedge small withdrawals) and transparent (to partially hedge large withdrawals), in order to mitigate liquidity risk to the maximum extent possible in the model. It shows that both liquidity and transparency are important dimensions of liquidity risk management, and may need to be combined to achieve a socially optimal outcome.

If the cost of hedging were higher, a bank could improve by foregoing the less efficient hedging mechanism, or avoiding hedging altogether. Yet, without loss of generality, we focus further analysis on possible distortions from optimal full hedging, and consider the case when conditions (1) or (2) are satisfied.

5 Suboptimal Risk Management

We now turn to bank's private liquidity and transparency choices. They can be distorted by leverage. The cost of hedging reduces bankers' payoff in the good state. At the same time, bankers do not fully internalize the benefits of lower probability of failure due to their limited liability (they share those benefits with debtholders or the deposit insurance fund). Consequently, when hedging choices are not contractible, banks can under-invest in hedging. (We introduce the possibility of contracting on hedging choices in the next section, in the context of regulation.) Such risk-shifting is a standard agency problem in corporate finance (Jensen and Meckling, 1976). For banks, similar distortions can also be derived from systemic externalities of bank failure (Acharya and Yorulmazer, 2007) or gambling for LOLR rents (Mailath and Mester, 1994, Ratnovski, 2007).

5.1 Private Payoffs

Consider the amount of debt the bank has to repay in case of success. At date 0, it borrowed 1 unit of money, with a nominal repayment 1 thanks to deposit insurance. When the bank refinances some debt at the intermediate date with new borrowing, that has zero net effect on debt outstanding (intermediate refinancing, when available, also has a 1 nominal interest rate: it is risk-free since provided only to banks known to be solvent). If a solvent bank repays L from the precautionary buffer at date 1, this reduces the debt outstanding to $1 - L$, which has to be repaid at date 2. As a result, the bank's total debt repayment in case of success is always 1. (Debt repayment in case of failure is 0.)

We can now derive the private payoffs. They are similar to the social ones, with the difference that the bank does not internalize the debt repayment of 1 that it makes in case of success. The payoffs for strategies "N", "L", "T" and "NT" are:

$$\begin{aligned}
\Pi_N &= (1-s-q)\cdot(X-1) \\
\Pi_L &= (1-s-q(1-l))\cdot(X-C-1) \\
\Pi_T &= (1-s-q(1-t))\cdot(X-C-1) \\
\Pi_{LT} &= (1-s-q(1-l)(1-t))\cdot(X-2C-1)
\end{aligned}$$

5.2 Risk Management Choices

Observe that in this model leverage does not affect the choice between liquidity and transparency: as in the social optimum, banks prefer liquidity for $l \geq t$ ($\Pi_L \geq \Pi_T$) and transparency for $l < t$ ($\Pi_L < \Pi_T$). The reason is that liquidity and transparency have the same cost, and bankers benefit from the effectiveness of the hedge they adopt.

We can now derive banks' private hedging choices. Recall that we focus on sufficiently low values of C ((1) for $l \geq t$ or (2) for $l < t$), such that it is socially optimal for banks to combine liquidity and transparency in risk management. As before, we distinguish two cases:

Case 1: Liquidity is more effective, $l \geq t$. The bank:

– Chooses to be only liquid ("L") or unhedged ("N") – deviating from the social optimum – for intermediate costs of hedging $\Pi_{LT} < \Pi_L$ (but restricted to $\Pi^S_{LT} \geq \Pi^S_L$):

$$\frac{q(1-l)t}{1-s-q(1-l)(1-2t)}\cdot(X-1) < C \leq \frac{q(1-l)t}{1-s-q(1-l)(1-2t)}\cdot X \tag{3}$$

– Chooses to be both liquid and transparent ("LT") – in line withy the social optimum – for low costs of hedging $\Pi_{LT} \geq \Pi_L$:

$$C \leq \frac{q(1-l)t}{1-s-q(1-l)(1-2t)}\cdot(X-1) \tag{4}$$

Case 2: Transparency is more effective, $l < t$. The bank:

– Chooses to be only transparent ("T") or unhedged ("N") for $\Pi_{LT} < \Pi_T$ (but restricted to $\Pi_{LT}^S \geq \Pi_T^S$):

$$\frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot (X-1) < C \leq \frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot X \qquad (5)$$

– Chooses to be both liquid and transparent ("LT") for $\Pi_{LT} \geq \Pi_T$:

$$C \leq \frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot (X-1)$$

Note that the bank is more likely to deviate from the social optimum for higher cost of hedging C and lower return in case of success (related to charter value) X. It is easy to observe from (3) and (5) that for any X there exist values of C such that public and private hedging incentives diverge, which leads to our next result:

Proposition 3 *A bank can under-invest in liquidity and transparency when its incentives are distorted by leverage. For any l, t, q, s, there exist values of C such that conditions (3) for $l \geq t$ or (5) for $l < t$ are satisfied.*

6 Multitasking in Liquidity Regulation

Banks' suboptimal hedging choices can justify regulatory intervention. Influencing the size of the bank's liquidity buffer is relatively easy, because the holdings of liquid assets are (to a large extent) verifiable and can be imposed, for example, by explicit ratios. However, as discussed in Section 2, regulatory lever on transparency is weaker and at best indirect.

This implementation issue (cf. Glaeser and Shleifer, 2001) can help explain why bank regulation typically puts emphasis on ensuring prudential liquidity buffers rather than transparency and market access. Yet when transparency is an important yet not verifiable component of risk management, the optimal design of liquidity regulation becomes a multi-tasking problem. The challenge is that liquidity requirements can affect banks' endogenous incentives to adopt transparency.

To see this analytically, consider the setting where:

- It is socially optimal that a bank is both liquid and transparent, $\Pi^S_{TL} \geq \Pi^S_T$ (2);
- Without regulation, the bank chooses to be transparent only: this implies $t > l$, and $\Pi_T > \Pi_{TL}$ and $\Pi_T \geq \Pi_N$:

$$\frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot (X-1) < C \leq \frac{qt}{1-s-q(1-t)} \cdot (X-1) \qquad (6)$$

Suppose now that authorities respond to suboptimal liquidity by imposing liquidity requirements. The aim is to restore socially optimal risk management, which *combines* liquidity and transparency. The problem is that, due to multitasking, this cannot always be achieved. In particular, there is a danger that, in response to liquidity requirements, a bank will stop investing in transparency.

Under liquidity requirements, the decision to retain transparency depends on its effectiveness as a second hedge. When transparency is very effective, compared to the effectiveness of liquidity and the cost of hedging, the bank is likely to preserve it on top of mandated liquidity. The bank would retain transparency for $\Pi_{LT} \geq \Pi_L$ (4) as determined by low C, low l, and high t. However when transparency is less effective, the bank can choose to drop transparency. This happens for $\Pi_{LT} < \Pi_L$:

$$C > \frac{q(1-l)t}{1-s-q(1-l)(1-2t)} \cdot (X-1) \qquad (7)$$

We show in the Appendix that there exist parameters value of X, l, t, q, s and C, such that the intersection of (2), (6) and (7) is nonempty:

$$\begin{aligned} C &\leq \min\left\{\frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot X \; ; \; \frac{qt}{1-s-q(1-t)} \cdot (X-1)\right\} \qquad (8) \\ C &> \max\left\{\frac{ql(1-t)}{1-s-q(1-2l)(1-t)} \cdot (X-1) \; ; \; \frac{q(1-l)t}{1-s-q(1-l)(1-2t)} \cdot (X-1)\right\} \end{aligned}$$

Then, under liquidity requirements, previously transparent banks lose incentives to invest in transparency and choose to remain with mandated liquidity only. Recall that in this setup transparency was a more effective method of hedging liquidity risk ($t > l$). Therefore when liquidity is substituted for transparency the probability of illiquidity-driven failures of solvent banks increases from $q(1-t)$ to $q(1-l)$, representing higher liquidity risks and lower social welfare.

Proposition 4 *Liquidity requirements reduce banks' incentives to invest in transparency. There exist values of parameters l, t, q, s and C which satisfy $t > l$ and (8), so that a bank stops investing in transparency in response to liquidity requirements, leading to higher risks and lower social welfare.*

Observe that transparency is likely to be effective ($t > l$) in countries with more developed financial markets. It is there that the adverse effects of incorrect liquidity requirements are most likely. In contrast, banks in developing countries have relatively limited market access ($t < l$) and can more safely emphasize stock liquidity. This is consistent with the evidence that liquidity regulation is not binding in advanced banking systems, such as those of US or UK (Bennet and Peristiani, 2002, Chaplin et al., 2000), while banks in developing countries often face stringent liquidity requirements (Freedman and Click, 2006). These cross-country differences in optimal liquidity-transparency outcomes may need to be born in mind during possible international convergence of liquidity regulation.

7 Discussion

Withdrawals and information The model assumed that the size of withdrawals (L with probability l and 1 with probability $1-l$) is independent of the concurrent informational signal. This reflects the behavior of uninformed depositors (who do not receive or cannot interpret the solvency signal) or term funding with pre-defined volumes of liabilities maturing at any point in time. Still, our model is robust to the possibility of high withdrawals being correlated with a signal of possible insolvency. Indeed, notice that in the model the size of withdrawals does not matter under a "positive" signal (a bank known to be solvent can refinance any withdrawals). Therefore, l and $1 - l$ can be interpreted as the probability of low and high withdrawals conditional on a "negative" signal. All results remain.

Deposit insurance For modelling convenience, we assumed that initial (date 0) funding is covered by non-contributory deposit insurance. This assured that the repayment bank had to promise initial investors and the bank's total net repayment over dates 1 and 2 were both always 1. Note that the deposit insurance assumption lowers the amount a bank has to repay in case of success, reducing risk-shifting and *increasing* hedging incentives. This works against our result of under-investment in hedging.

The model is therefore robust to altering the deposit insurance set-up, for example by considering no deposit insurance or fairly priced deposit insurance. Then the ultimate cost of insufficient hedging would be born by bankers themselves, not the public deposit insurance fund. Yet when liquidity and transparency choices are not contractible, and deposit insurance premia are set *before* banks make hedging choices, banks will still under-insure in equilibrium, distorting social welfare. The multi-tasking problem in liquidity regulation will also remain.

Other effects of liquidity and transparency Banks' asset buffers and transparency may affect some other dimensions of bank performance. For example, liquidity can provide bankers with private benefits of control (Myers and Rajan, 1998) or allow them to conceal losses (Rajan, 1994). Transparency can improve bank performance when it increases fundamental value and reduces overall funding costs (for example by facilitating screening and monitoring: Boot and Schmeits, 2000, Flannery 2001), or have a negative impact when information is passed to competitors. Those effects are outside the scope of this model. This paper has a specific focus on the effects of liquidity and transparency during sudden shocks. The reason is that such shocks are particularly likely to be associated with increased solvency uncertainty – the key driver of bank liquidity risk.

Systemic liquidity events The model of liquidity risk developed in this paper explicitly focuses on idiosyncratic liquidity events. We assumed the presence of outside investors willing to lend to banks as long as they can obtain appropriate return. Such description of bank-specific refinancing frictions applies well to a number of recent episodes in developed countries, the most recent ones being *Lehman Brothers* in 1998 and *Commerzbank* in 2002.

Yet the credit turmoil of 2007– was characterized by an *aggregate* shortage of liquidity in the banking system. The apparent reason was that informed potential lenders were themselves exposed to liquidity needs (see Brunnermeier and Pedersen, 2008, or Caballero and Krishnamurthy, 2008), while outsiders with spare liquidity were prevented from lending by informational frictions (see e.g. Acharya and Yorulmazer, 2008). Our model offers insight into the mechanics of such events. If example, it can be applied to help explain sudden shifts in lending capacity of outside investors. The model demonstrates why in normal condition it is easy for the banking system as a whole to raise wholesale funds (e.g. from non-financial corporations, high net worth individuals, or foreign entities), and why such lending to the banking system can be subject to abrupt stops and flight to quality.

Empirical implications Empirical literature showed that both stock liquidity (Paravisini, 2007) and transparency (better access to external financing, Kashyap and Stein, 1990, Holod and Peek, 2004) determine bank financial constraints. There is evidence that banks may be insufficiently liquid (Gatev et al., 2004, Gonzalez-Eiras, 2003) or transparent (Morgan, 2002). Yet the literature has mostly considered liquidity and market access separately[6]. Our paper articulates a descriptive joint framework, emphasizing a certain but limited in size hedging effect of cash holdings, and a more flexible in size but uncertain hedging effect of transparency and market access. This offers a

[6] A rare exception is Acharya et al. (2007) who show that debt capacity is a more effective way of boosting investment in future high cash flow states, while retained earnings – in low cash flow states.

number of novel empirical predictions for bank liquidity risk and its management.

Firstly, the model indicates that the correct measurement of bank financial constraints should include both its liquidity position and ability to borrow. Secondly, the mix liquidity insurance choices affects banks' resilience to shocks of different magnitude. More liquid banks have higher resilience to small shocks, while more transparent banks can better withstand large shocks. Lastly, the model has predictions for banks' choice between liquidity and transparency. Other else equal, more liquid banks will choose to be less transparent, while more transparent banks (for example, listed ones) – less liquid. Also, banks will rely on liquidity to manage routine cash flows, but emphasize transparency (borrowing capacity) in the anticipation of large liquidity needs.

8 Conclusions

This paper modelled liquidity risk driven by imperfect information in "informed" wholesale funding markets. We argued that the primary driver of such risk are solvency concerns that can arise at the refinancing stage. The setting offered a number of risk management and regulatory implications. We showed that banks can combine liquidity buffers and transparency (enhanced ability to borrow) in risk management. Yet their private choices can be distorted by leverage, while regulation complicated by multi-tasking. Liquidity requirements can lead to reduced banks' transparency and under some conditions increase liquidity risk. We also argued that solvency regulation cannot fully address liquidity risks due to remaining information imperfections, and that liquidity risks are likely to persist as banks shift from retail deposits to wholesale funding.

A Appendix

Proof that condition (8) is nonempty Recall (8):

$$\begin{aligned}
C &< \min\left\{\frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot X\ ;\ \frac{qt}{1-s-q(1-t)}\cdot(X-1)\right\}\\
C &> \max\left\{\frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot(X-1)\ ;\ \frac{q(1-l)t}{1-s-q(1-l)(1-2t)}\cdot(X-1)\right\}
\end{aligned}$$

Consider the fist inequality. Observe that for $X > 2$

$$\frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot X < \frac{qt}{1-s-q(1-t)}\cdot(X-1)$$

because in the nominator, $ql(1-t)X < qt(X-1)$ ($X > 2$ and $t > l > l(1-t)$), and in the denominator, $q(1-t) > q(1-2l)(1-t)$.

Therefore for $X > 2$ (8) transforms into

$$\begin{aligned}
C &< \frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot X\\
C &> \max\left\{\frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot(X-1)\ ;\ \frac{q(1-l)t}{1-s-q(1-l)(1-2t)}\cdot(X-1)\right\}
\end{aligned}$$

Observe that there always exist C such that

$$\frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot(X-1) < C < \frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot X$$

and there exist C such that

$$\frac{q(1-l)t}{1-s-q(1-l)(1-2t)}\cdot(X-1) < C < \frac{ql(1-t)}{1-s-q(1-2l)(1-t)}\cdot X$$

at least for t close to but above l since the two fractions become identical and $X - 1 < X$. *QED.*

References

[1] Acharya, Viral V., Heitor Almeida and Murillo Campello (2007) "Is Cash Negative Debt? A Hedging Perspective on Corporate Financial Policies", *Journal of Financial Intermediation*, 16(4): 515-54

[2] Acharya, Viral V. and Tanju Yorulmazer (2007) "Too Many to Fail – An Analysis of Time-inconsistency in Bank Closure Policies". *Journal of Financial Intermediation* 16, 1-31

[3] Acharya, Viral V. and Tanju Yorulmazer (2008) "Cash-in-the-Market Pricing and Optimal Resolution of Bank Failures", *Review of Financial Studies*, forthcoming

[4] Admati, Anat, and Paul Pfleiderer (2000) "Forcing Firms to Talk: Financial Disclosure Regulation and Externalities", *Review of Financial Studies* 13(3): 479-519

[5] Bennet, Paul and Stavros Peristiani (2002) "Are US Reserve Requirements still Binding?", *FRBNY Economic Policy Review* 8(1)

[6] Boot, Arnoud and Anjolein Schmeits (2000) "Market Discipline and Incentive Problems in Conglomerate Firms with Applications to Banking", *Journal of Financial Intermediation*, 9(3) 1-34.

[7] Boot, Arnoud, and Anjan Thakor (2000) "Can Relationship Banking Survive Competition?", *Journal of Finance*, 55(2): 679-713.

[8] Boot, Arnoud, and Anjan Thakor (2001) "The Many Faces of Information Disclosure", *Review of Financial Studies*, 14(4): 1021-1057

[9] Brunnermeier, Markus and Lasse Pedersen (2008) "Market Liquidity and Funding Liquidity", *Review of Financial Studies*, forthcoming

[10] Caballero, Ricardo and Arvind Krishnamurthy (2008) "Collective Risk Management in a Flight to Quality Episode", *Journal of Finance*, forthcoming

[11] Calomiris, Charles W. (1999) "Building an Incentive Compatible Safety Net", *Journal of Banking and Finance*, 23: 1499-1519

[12] Chaplin, Graeme, Alison Emblow and Ian Michael, 2000, "Banking System Liquidity: Developments and Issues", *Financial Stability Review,* Dec: 93-112

[13] Chari, V.V. and R. Jagannathan (1988), "Banking Panics, Information, and Rational Expectations Equilibrium", *Journal of Finance* 43: 749-61.

[14] Diamond, Douglas W., Philip H. Dybvig (1983) "Bank Runs, Deposit Insurance, and Liquidity", *Journal of Political Economy* 91(3): 401-419

[15] Diamond, Douglas W., Raghuram G. Rajan (2001) "Liquidity Risk, Liquidity Creation, and Financial Fragility: A Theory of Banking," *Journal of Political Economy*, 109(2): 287-327

[16] Flannery, Mark (1996) "Financial Crises, Payment System Problems, and Discount Window Lending," Journal of Money, Credit and Banking, 28(4): 804-24

[17] Flannery, Mark (2001) "The Faces of Market Discipline" *Journal of Financial Services Research* 107-119.

[18] Freedman, Paul and Reid Click (2006) "Banks that Don't Lend? Unlocking Credit to Spur Growth in Developing Countries", *Development Policy Review*, 24 (3), 279-302.

[19] Freixas, Xavier, Bruno M Parigi and Jean-Charles Rochet (2004) "The Lender of the Last Resort: A Twenty-First Century Approach", *Journal of the European Economic Association* 2(6)

[20] Gatev, Evan and Philip E. Strahan (2006) "Banks' Advantage in Hedging Liquidity Risk: Theory and Evidence from the Commercial Paper Market" *Journal of Finance* 61(2): 867-92

[21] Gatev, Evan, Til Schuermann, Philip E. Strahan (2004) "How do Banks Manage Liquidity Risk? Evidence from Equity and Deposit Markets in the Fall of 1998", NBER working paper 10982

[22] Glaeser, Edward L. and Andrei Shleifer (2001) "A Reason for Quantity Regulation", *American Economic Review* 91(2), : 431-435.

[23] Griffin, Paul K. and Samoa Wallach (1991) "Latin American Lending by Major U.S. Banks: The Effects of Disclosures about Nonaccrual Loans and Loan Loss Provisions", *Accounting Review*, 86(4): 830-846

[24] Gonzalez-Eiras, Martin (2004) "Banks' Liquidity Demand in the Presence of a Lender of Last Resort". working paper Universidad de San Andres, Buenos Aires

[25] Goodfriend, Marvin and Robert G. King (1998) "Financial Deregulation, Monetary Policy, and Central Banking", *Federal Reserve Bank of Richmond Economic Review*, May/June: 3-22

[26] Holod, Dmytro and Joe Peek (2005) "Asymmetric Information and Liquidity Constraints: A New Test", FDIC Center for Financial Research Working Paper No. 2005-02

[27] Jensen, Michael C. and William Meckling (1976) "Theory of the Firm: Managerial Behavior, Agency Costs and Ownership Structure", *Journal of Financial Economics*

[28] Kashyap, Anyl K. and Jeremy C. Stein (2000), "What do a Million Observations on Banks Say about the Transmission of Monetary Policy?" *American Economic Review*, 90(3): 407-28.

[29] Mailath, George J. and Loretta J. Mester (1994) "A Positive Analysis of Bank Closure" *Journal of Financial Intermediation*, 3(3) 272-299

[30] Morgan, Donald P. (2002), "Rating Banks: Risks and Uncertainty in an Opaque Industry", *American Economic Review* 92: 874-88.

[31] Myers, Stewart C. and Raghuram G. Rajan (1998) "The Paradox of Liquidity", *Quarterly Journal of Economics*, 113(3): 733-771

[32] Paravisini, Daniel (2007), "Local Bank Financial Constraints and Firm Access to External Finance", *Journal of Finance*, forthcoming

[33] Perotti, Enrico C. and Ernst-Ludwig von Thadden (2003) "Strategic Transparency and Informed Trading: Will Globalization Force Convergence of Corporate Governance? ", *Journal of Financial and Quantitative Analysis*, 38(1): 61-85.

[34] Rajan, Raghuram G. (1994) "Why Bank Credit Policies Fluctuate: A Theory and Some Evidence" *Quarterly Journal of Economics* 109(2): 399-441

[35] Ratnovski, Lev (2007) "Bank Liquidity Regulation and the Lender of Last Resort", *Journal of Financial Intermediation*, forthcoming

[36] Rochet, Jean-Charles and Xavier Vives (2004) ""Coordination Failure and the Lender of Last Resort: Was Bagehot Right After All?", *Journal of the European Economic Association*, 2(6): 1116-47.

FIGURE 1

<u>The timeline</u>

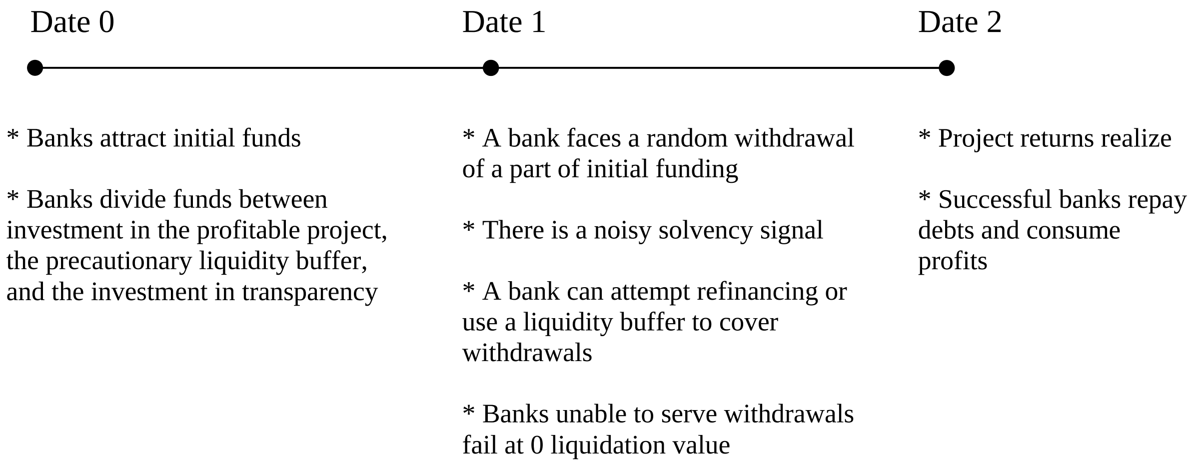


FIGURE 2

<u>Information structure at the Intermediate date</u>

<table>
<tr><td>Fundamentals</td><td colspan="2">Solvent, 1–s</td><td>Insolvent, s</td></tr>
<tr><td>Information at date 1</td><td>Positive signal, known solvent</td><td colspan="2">Negative signal, pooled together</td></tr>
<tr><td>Outcome</td><td>1–s–q

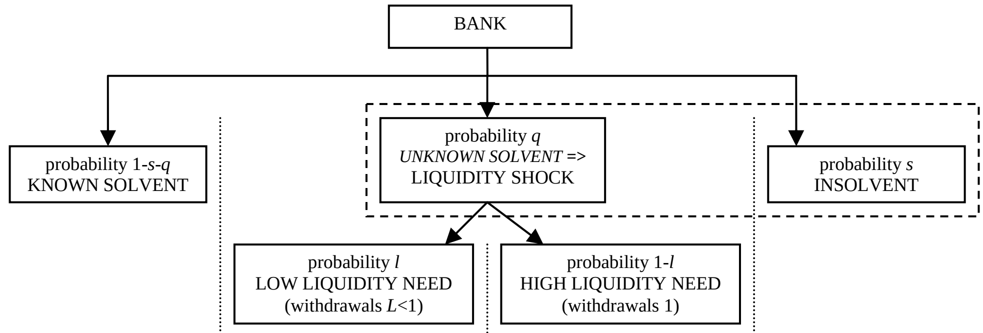
Able to refinance</td><td>q

Solvent, but unable to refinance</td><td>s

Insolvent</td></tr>
</table>

FIGURE 3

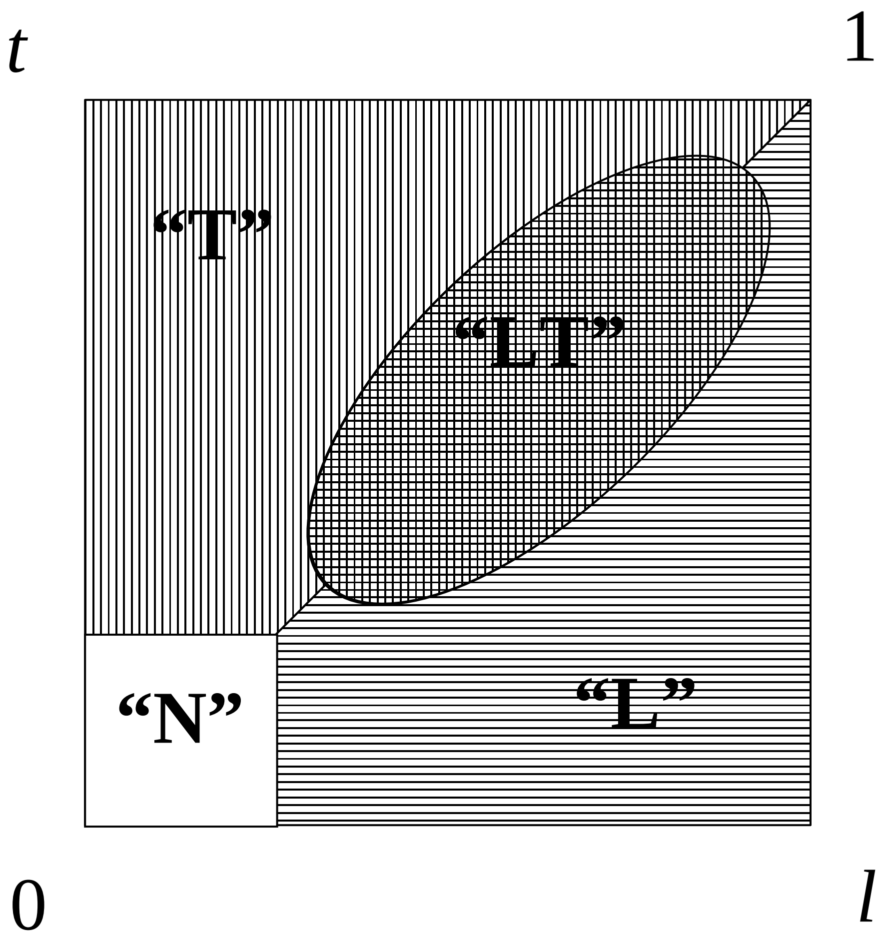
<u>Bank returns, for states at date 1, depending on ex-ante hedging decisions</u>



EX-ANTE HEDGING DECISIONS	probability 1-*s*-*q* KNOWN SOLVENT	probability *l* LOW LIQUIDITY NEED (withdrawals *L*<1)	probability 1-*l* HIGH LIQUIDITY NEED (withdrawals 1)	probability *s* INSOLVENT
NO HEDGE	Survives, returns *X*	Fails, returns *0*	Fails, returns *0*	Fails, returns *0*
LIQUID (can refinance small shocks)	Survives, returns *X-C*	Survives, returns *X-C*	Fails, returns *0*	Fails, returns *0*
TRANSPARENT (can refinance any shoks w.p. *t*)	Survives, returns *X-C*	Survives w/p *t*, returns *X-C* Fails w/p *1-t*, returns *0*	Survives w/p *t*, returns *X-C* Fails w/p *1-t*, returns *0*	Fails, returns *0*
BOTH	Survives, returns *X-2C*	Survives, returns *X-2C*	Survives w/p *t*, returns *X-2C* Fails w/p *1-t*, returns *0*	Fails, returns *0*

FIGURE 4

Hedging equilibria



Financial Innovation, Macroeconomic Stability and Systemic Crises*

Prasanna Gai
Australian National University and Bank of England

Sujit Kapadia†
Bank of England

Stephen Millard
Bank of England

Ander Perez
London School of Economics

October 2007‡

Abstract

We present a general equilibrium model of intermediation designed to capture some of the key features of the modern financial system. The model incorporates financial constraints and state-contingent contracts, and contains a clearly defined pecuniary externality associated with asset fire sales during periods of stress. If a sufficiently severe shock occurs during a credit expansion, this externality is capable of generating a systemic financial crisis that may be self-fulfilling. Our model suggests that financial innovation and greater macroeconomic stability may have made financial crises in developed countries less likely than in the past, but potentially more severe.

Keywords: Systemic Financial Crises; Financial Innovation; Macroeconomic Stability; Modern Financial Systems; Fire Sales.
JEL Classification Numbers: E32, E44, G13, G2.

---

*This paper represents the views of the authors and should not be thought to represent those of the Bank of England or Monetary Policy Committee members. We are grateful to Jagjit Chadha, John Eatwell, Andrew Glyn, Andy Haldane, Roman Inderst, Nigel Jenkinson, Karsten Jeske, Nobu Kiyotaki, Arvind Krishnamurthy, Bill Nelson, Tanju Yorulmazer, two anonymous referees, and seminar participants at the Bank of England, the London School of Economics, the Federal Reserve Bank of Atlanta conference on "Modern Financial Institutions, Financial Markets, and Systemic Risk", the CERF conference on "The Changing Nature of the Financial System and Systemic Risk", the 2006 European and North American Summer Meetings of the Econometric Society, the Federal Reserve Bank of San Francisco conference on "Financial Innovations and the Real Economy", the 2007 RES conference, and the FMG conference on "Cycles, Contagion and Crises" for helpful comments and suggestions. Ander Perez gratefully acknowledges financial support from the Fundacion Rafael del Pino.

†Corresponding author. Bank of England, Threadneedle Street, London, EC2R 8AH, United Kingdom. E-mail: sujit.kapadia@bankofengland.co.uk

"When [financial] innovation ... takes place in a period of generally favorable economic and financial conditions, we are necessarily left with more uncertainty about how exposures will evolve and markets will function in less favorable circumstances. The past several years of exceptionally rapid growth in credit derivatives and the larger role played by non-bank financial institutions, including hedge funds, has occurred in a context of ... relatively strong and significantly more stable economic growth, less concern about the level and volatility in future inflation, and low expected volatility in many asset prices. Even if a substantial part of these changes prove durable, we know less about how these markets will function in conditions of stress..." (Geithner, 2006)

Systemic financial crises often occur when investment booms and rapid credit expansions collapse because the expectations of high future returns that drove them are not fulfilled (Borio and Lowe, 2002; Eichengreen and Mitchener, 2003). But while investment booms and busts have been an important part of recent financial crises in emerging market economies, their impact on financial stability in the advanced economies has been less marked. Greater macroeconomic stability and the growing sophistication of financial intermediation appear to have reduced the incidence of crisis. Increasingly, however, policymakers have become concerned that while these factors may have helped to reduce the likelihood of systemic crises, their impact, should one occur, could be on a significantly larger scale than hitherto (see, for example, Rajan, 2005, Tucker, 2005, and Gieve, 2006).[1]

It is difficult to make judgments on such issues without formally modelling the underlying externalities associated with systemic financial crises. One strand of the literature (e.g. Aghion *et al.*, 1999; Aghion *et al.*, 2001) draws on Kiyotaki and Moore (1997) to highlight credit frictions arising from enforcement problems.[2] These papers illustrate how endogenous balance sheet constraints, and financial development more generally,

[1] Gai *et al.* (2007) discuss the implications of these issues for risk assessment work at the Bank of England.

[2] An alternative strand of the literature highlights coordination problems amongst financial market participants as the key externality driving financial crises. See, for example, Diamond and Dybvig (1983), Obstfeld (1996), and Morris and Shin (1998).

contribute to financial instability. But since these papers do not permit state-contingent financial contracts, the extent to which the underlying externality drives their results is unclear. By contrast, in existing models with state-contingent contracts (e.g. Kehoe and Levine, 1993; Krishnamurthy, 2003; Lorenzoni, 2005; Gai *et al.*, 2006), investment projects are never abandoned and crises never occur. Moreover, these papers do not consider the effects of financial innovation or changes in macroeconomic volatility.

This paper seeks to bridge this gap. We develop a general equilibrium model of intermediation with financial constraints and state-contingent contracts. Systemic financial crises are generated through a clearly defined pecuniary externality associated with asset 'fire sales' during periods of stress. Moreover, the potential for instability is present *ex ante* and does not rely on sunspots or other undefined factors external to the model.

In our setup, consumers channel funds through collateral-constrained financial intermediaries to firms operating in more-productive sectors of the economy. Firms manage investment projects but intermediaries retain financial control over them. Even though financial contracts can be made contingent on the aggregate state, enforcement problems mean that insurance opportunities for intermediaries are limited. As a result, adverse aggregate shocks to the productive sectors of the economy may force intermediaries to sell capital to less-productive sectors to remain solvent. In the spirit of Fisher (1933) and Shleifer and Vishny (1992), this distress selling is associated with reduced asset prices.[3] In turn, this creates a feedback to net worth which affects the balance sheets of all intermediaries, potentially leading to further asset sales. Since intermediaries do not internalise the effect on asset prices of their own sales, the competitive equilibrium is constrained inefficient. In extreme cases, it is this externality which can result in a systemic financial crisis that may be self-fulfilling.

The analysis points to a range of possible outcomes. Since expected future returns in productive sectors are high, initial investment is always strong and associated with a large credit expansion. Provided that there is *no adverse shock*, investment and credit growth remain robust, and there are no asset sales. For *mild negative shocks*, firms

[3] In a study of commercial aircraft transactions, Pulvino (1998) finds evidence for this type of fire sale effect; Coval and Stafford's (2007) analysis of mutual fund asset sales demonstrates that these effects may be present even in highly liquid markets.

and intermediaries liquidate some of their assets. However, since intermediaries remain solvent and firms continue to operate in productive sectors, this outcome can be viewed as a 'recession' rather than a systemic crisis.

For *more severe shocks*, multiple equilibria can arise, with (*ex ante*) beliefs determining the actual equilibrium which results. Multiplicity can occur in bad states because the supply of capital by intermediaries during fire sales is downward sloping in price, since the lower the price, the more capital they will have to sell to remain solvent. If agents have 'optimistic' beliefs about how the economy will evolve under stress, there will only be a partial liquidation of assets, as in the 'recession' case. But if beliefs are 'pessimistic', a systemic financial crisis occurs. Moreover, for *extremely severe shocks*, a crisis is inevitable, regardless of beliefs. Under this scenario, asset prices are driven down to such an extent that all intermediaries and firms are forced to liquidate all of their assets – a full-blown financial crisis occurs, intermediaries shut down, and the closure of firms means that there are no investment opportunities in the more-productive sectors of the economy.

The financial system has been changing rapidly in recent years. Intermediation is increasingly conducted through non-bank intermediaries such as private equity firms and hedge funds, who typically have higher leverage in risk-adjusted terms than traditional banks. Resale markets for capital have deepened, and sophisticated financial products and contracts, such as credit derivatives and asset-backed securities, have mushroomed (White, 2004; Allen and Gale, 2007; Plantin *et al.*, 2007). Our model suggests that these developments may have made economies less vulnerable to crises as they widen access to liquidity and allow assets to be traded more easily during periods of stress. But, by relaxing financial constraints facing borrowers, they imply that, should a crisis occur, its impact could be more severe than previously.

We demonstrate how these effects may be reinforced by greater macroeconomic stability.[4] Our model predicts that mean preserving reductions in volatility make crises less likely since severe shocks occur less frequently. However, greater stability also

[4] A range of empirical studies (e.g. Benati, 2004; Stock and Watson, 2005) find that output and inflation volatility have fallen in many developed countries in recent years.

makes 'recession' states less likely. As a result, consumers are more willing to lend, allowing intermediaries to increase their borrowing and initial investment. But, if a crisis does then ensue, losses will be greater. Overall, our findings thus make clear how financial innovation and increased macroeconomic stability may serve to reduce the likelihood of crises in developed countries, but increase their potential impact.

Our paper has several points of contact with the literature. The model has some similarities to Holmstrom and Tirole (1998) and Jermann and Quadrini (2006), and builds on Lorenzoni's (2005) analysis of lending under endogenous financial constraints and asset prices. It differs in two key respects. First, we show how multiple equilibria and systemic crises can arise in such a model. Second, we capture some of the key features of intermediation in the modern financial system: though our model also applies to traditional banks, it is especially relevant to the activities of hedge funds, private equity firms, and other non-bank financial institutions. These developments allow us to model the effects of financial innovation and greater macroeconomic stability on the likelihood and potential scale of systemic crises.

In recent work, Allen and Carletti (2006) also assess the systemic effects of financial innovation. But they have a specific focus on credit risk transfer between banks and insurance companies, and on how its effects differ according to the type of liquidity risk that banks face. In particular, their model highlights how, in some circumstances, credit risk transfer can create the potential for contagion from the insurance sector to the banking sector, and thus be detrimental. By contrast, we consider the more general consequences of financial innovation through its broader impact on financial constraints and the depth of resale markets.[5]

The rest of the paper is structured as follows. Section 1 presents the basic structure of the model, while section 2 solves for equilibrium and discusses how multiplicity and systemic financial crises arise. Section 3 considers the effects of financial innovation and changes in macroeconomic volatility on the likelihood and potential scale of financial crises. A final section concludes.

[5] Financial innovation may also increase uncertainty about the behaviour of financial markets. We leave this issue aside and just focus on capturing the effects of certain trends linked to financial innovation.

1 The Model

The economy evolves over three periods ($t = 0, 1, 2$) and has two goods, a consumption good and a capital good. Consumption goods can always be transformed one for one into capital goods, but not vice versa. Because of the irreversibility of investment, the price of the capital good in terms of the consumption good (the asset price), q, may be less than one in the event of asset sales – this is one of the key drivers of our results.

1.1 *Financial Intermediaries and Other Agents*

The economy is composed of consumers, financial intermediaries, and firms, with large numbers of each type of agent. All agents are risk-neutral and identical within their grouping, and there is no discounting.

Consumers aim to maximise total consumption, $c_0 + c_1 + c_2$, where c_t is consumption in period t. They each receive a large endowment, e, of the consumption good in every period. Since they are only able to produce using a relatively unproductive technology operating in the traditional sector of the economy, they channel funds through intermediaries to firms operating in the more-productive sector of the economy.[6]

Intermediaries in the model are best viewed as operating in the modern financial system: they could be interpreted as traditional banks, but our model is also designed to apply to the activities of hedge funds, private equity firms, and other non-bank financial institutions. They borrow from consumers and invest in firms in order to maximise total profits, $\pi_0 + \pi_1 + \pi_2$, where profits and consumption goods are assumed to be interchangeable. However, their wealth is relatively limited: although they receive an endowment, n_0, of the consumption good in period 0 (this may be thought of as their initial net worth), this is assumed to be very small relative to e. We also assume that intermediaries are unable to trade each other's equity due to limited commitment, though

[6]Although intermediaries clearly have an important role in practice, there is nothing in the structure of our model which precludes consumers from investing directly in firms. We could formally motivate the existence of intermediaries by, for example, introducing asymmetric information or, more specifically, following Diamond and Dybvig (1983) or Holmstrom and Tirole (1998). But this would significantly complicate the analysis without changing our main results. Therefore, for simplicity and transparency, we simply assume that consumers can only invest in the more-productive sector through intermediaries. Indeed, the involvement of intermediaries in investment projects in the more-productive sector could be interpreted as partially driving the higher returns in that sector relative to the traditional sector.

relaxing this assumption does not affect our qualitative results.

Firms have no special role in our setup. They are agents with no net worth who manage investment projects in exchange for a negligible payment – this could be viewed as following from perfect competition amongst firms. Since this implies that intermediaries effectively have complete control over investment projects, we abstract from the behaviour of firms in all of what follows, and simply view intermediaries as having direct access to the productive technology.

The assumption that intermediaries have financial control over firms may appear somewhat extreme. But it embeds some of the recent developments in financial markets in a simple way. In particular, as Plantin *et al.* (2007) stress, the greater use of sophisticated financial products such as credit derivatives, and the deepening of resale markets for capital have made it easier for intermediaries to trade their assets (i.e. their loans / investments in firms). This especially applies to non-traditional financial intermediaries.

1.2 *Production Opportunities*

Figure 1 depicts the timing of events. Intermediaries can invest in the productive sector in periods 0 and 1. Since there is no depreciation, an investment of i_0 in period 0 delivers i_0 units of capital in period 1. We also suppose it delivers xi_0 units of the consumption good (profit) in period 1, where x is a common aggregate shock with distribution function $H(x)$. The realisation of x is revealed to all agents in period 1, depends on the aggregate state, s, and can be contracted upon. Intuitively, the shock represents the per unit surplus (positive x) or shortfall (negative x) in period 1 revenue relative to (future) operating expenses. Alternatively, a positive x could be viewed as an early return on investment and a negative x as a restructuring cost or an additional capital cost which must be paid to continue with the project. Under both interpretations, a negative x does not need to be paid by anyone if the investment project is abandoned. But, when analysing the welfare gains associated with the social planner's solution, we allow for the possibility that an unpaid negative x imposes a cost to society of $w = -\lambda x$, where $0 < \lambda < 1$.

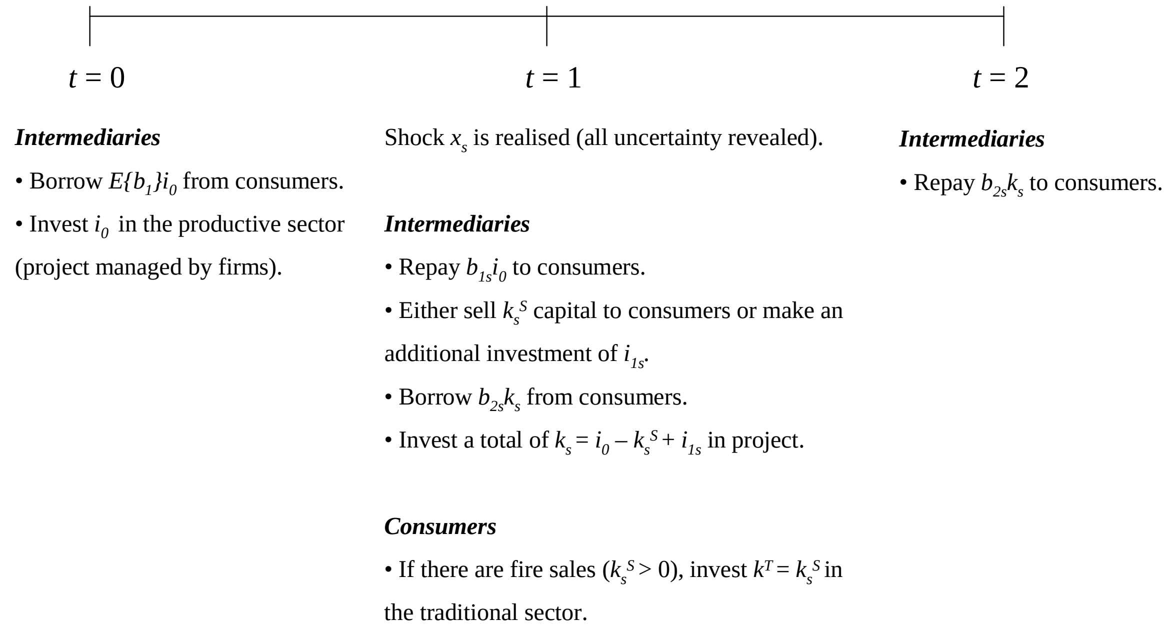


Figure 1: Timeline of Events

Let $E(x) = \mu > 0$, so that early investment in period 0 is expected to be profitable. If x turns out to be negative, the intermediary has two options: it can either incur the cost xi_0 (possibly by selling a portion of its capital to consumers) and continue with the investment project; or it can go into liquidation, abandoning the project and selling all of its capital to consumers.[7] In the latter case, it receives zero profit in period 2 but does not need to pay xi_0. In what follows, we associate total liquidation by the representative intermediary as reflecting a systemic financial crisis.[8]

In period 1, intermediaries can either sell k^S units of capital to consumers or make an additional investment, $i_1 \geq 0$. Therefore, they enter period 2 owning a total capital stock of:

$$k_s = i_0 - k_s^S + i_{1s}. \tag{1}$$

Invested in the productive sector, this capital yields Ak_s units of the consumption good in period 2, where A is a constant greater than one.

[7]Since intermediaries are homogeneous and unable to trade each other's equity, there is no scope for them to sell capital to each other following a negative aggregate shock.

[8]As financial contracts are fully state-contingent in this model (see section 1.3), they will be specified so that repayments from intermediaries to consumers are zero in states in which intermediaries are solvent but in severe distress. Since this implies that intermediaries never default on their contractual liabilities to consumers, it makes sense to associate systemic financial crises with total liquidation.

If consumers acquire capital from intermediaries in period 1, they can also use it to produce consumption goods in period 2, but they only have access to a less-productive technology operating in the perfectly competitive traditional sector of the economy. In particular, the production function in the traditional sector, $F\left(k^T\right)$, displays decreasing returns to scale, with $F'\left(k^T\right) > 0$ and $F''\left(k^T\right) < 0$. For simplicity, $F'(0) = 1$, implying that there is no production in the traditional sector unless $q < 1$ (i.e. unless intermediaries sell capital in period 1). To aid intuition, we assume the specific form:

$$F\left(k^T\right) = k^T\left(1 - \alpha k^T\right), \tag{2}$$

where $2\alpha k^T < 1$. We also assume that capital used in the traditional sector depreciates fully after one period, so that it is worthless in period 2.

The diminishing returns embedded in the production function are designed to capture the link, highlighted by Shleifer and Vishny (1992), between distress selling of capital and reduced asset prices. As they argue, many physical assets (e.g. oil tankers, aircraft, copper mines, laboratory equipment etc.) are not easily redeployable, and the portfolios of intermediaries, many of which contain exotic tailor-made assets, are similar in this regard. Therefore, if an aggregate shock hits an entire sector, participants in that sector wishing to sell assets may be forced to do so at a substantial discount to industry outsiders.

The parameter α reflects the productivity of second-hand capital. Although this partly depends on the underlying productivity of capital in alternative sectors, it also captures the effectiveness with which capital is channelled into its most effective use when it is sold. As such, it is likely to be decreasing in financial market depth (note that $\alpha = 0$ corresponds to constant returns to scale in the traditional sector). Since increased market participation, greater global mobility of capital, and the development of sophisticated financial products may all serve to deepen resale markets, α is likely to have fallen in recent years.

1.3 *Financial Contracts and Constraints*

Intermediaries partially finance investment projects by borrowing. At date 0, they offer a state-contingent financial contract to consumers. As shown in the timeline, this specifies repayments in state s of $b_{1s}i_0$ in period 1 and $b_{2s}k_s$ in period 2, and borrowing of $E(b_1)\,i_0$ in period 0 and $b_{2s}k_s$ in period 1 and state s, where b is the repayment / borrowing ratio. Since period 1 repayments to consumers on period 0 lending are state-contingent, this has some features of an equity contract. In particular, the contract is capable of providing intermediaries with some insurance against aggregate shocks.

Although this contract is fully contingent on the aggregate state, it is subject to limited commitment and potential default. This friction is fundamental to the model: without it, the competitive equilibrium would be efficient and systemic financial crises would never occur. Its significance lies in the borrowing constraints which it imposes on financial contracts:

$$(b_{1s}i_0 - b_{2s}k_s) + b_{2s}k_s \geq 0 \qquad \forall s, \tag{3}$$

$$b_{2s}k_s \geq 0 \qquad \forall s, \tag{4}$$

$$b_{1s}i_0 \leq \theta q_{1s}i_0 \qquad \forall s, \tag{5}$$

$$b_{2s}k_s \leq \theta q_{2s}k_s \qquad \forall s, \tag{6}$$

where q_{ts} is the asset price in period t and state s, and $\theta \leq 1$ is the fraction of the asset value that can be used as collateral.

The first two constraints, (3) and (4), reflect limited commitment on the consumer side. In particular, they imply that net future repayments to consumers must be non-negative. In other words, regardless of the state, consumers cannot commit to make net positive transfers to intermediaries at future dates. Constraint (3) relates to net future repayments as viewed in period 0 (for which additional intermediary borrowing in period 1 must be taken into account); constraint (4) relates to future repayments as viewed in period 1. These constraints follow from assuming that the future income of consumers cannot be seized – consumers can always default on their financial obligations.[9]

[9]Collectively, it would be in the interests of consumers to commit to make net positive transfers to

The final two constraints, (5) and (6), specify that intermediaries can only borrow up to a fraction, θ, of the value of their assets in each period, where we define θ to be the maximum loan-to-value ratio. Jermann and Quadrini (2006, Appendix B) present a simple model which motivates constraints such as these. In particular, they link an equivalent parameter to θ to the value of capital recovered upon default relative to its original value when held by the borrower, and to the relative bargaining power of borrowers and lenders. Importantly, if the recovery rate is less than one, the maximum loan-to-value ratio will also be less than one. As argued by Gai *et al.* (2006), recovery rates below one may reflect transaction costs built into the specifics of collateral arrangements, such as dispute resolution procedures. Alternatively, there may be human capital loss associated with default.

We regard the maximum loan-to-value ratio as being linked to the level of financial market development. It seems likely that financial innovation may have increased θ in recent years. Deeper resale markets may have reduced the human capital loss associated with default, and could have enabled sellers of assets seized upon default to pass on a larger proportion of the resale transaction costs to buyers than previously.[10] More generally, the greater use of credit derivative and syndicated loan markets may have increased recovery rates for lenders. Alternatively, as highlighted by Jermann and Quadrini (2006), the development of more sophisticated asset-backed securities may have made it easier for borrowers to pledge their assets as collateral to lenders. All of these factors may have made investors willing to accept higher loan-to-value ratios, thus raising θ.

It is clear that some of these factors relate to the depth of secondary markets. As such, increases in θ may be closely tied to reductions in α. This concurs with broader theoretical arguments linking the debt capacity of investors to the liquidity and depth of the secondary markets for assets used as collateral for that debt. For example, Williamson (1988) and Shleifer and Vishny (1992) discuss how the redeployability of assets is a key

intermediaries in certain states at future dates. But such a commitment is not incentive compatible since consumers each have an individual incentive to renege *ex post*. Limited commitment on the consumer side can thus also be viewed as stemming from the lack of a suitable commitment device amongst consumers.

[10] The latter point could potentially be modelled formally in a Nash bargaining framework – for a related model in this spirit, see Duffie *et al.* (2005).

factor in determining their liquidation value and that this, in turn, affects investors' debt capacity. More recently, Brunnermeier and Pedersen (2006) have studied the relationship between the leverage capacity of traders and financial market liquidity, demonstrating that they are likely to be positively correlated and, importantly, that causality can run both ways.

2 Equilibrium

We now solve for equilibrium, focusing primarily on the competitive outcome. Since consumers expect investment in the productive sector of the economy to be profitable, and since they have very large endowments relative to financial intermediaries, they always meet the borrowing demands of intermediaries provided that constraints (3)-(6) are satisfied. Meanwhile, as noted above, firms simply manage investment projects for a negligible wage. Therefore, we can solve for the competitive equilibrium by considering the optimisation problem of the representative intermediary.

2.1 *The Representative Intermediary's Optimisation Problem*

The representative intermediary's optimisation problem is given by:

$$\max_{\pi_0, \{\pi_{1s}\}, i_0, \{k_s\}, \{b_{1s}\}, \{b_{2s}\}} E_0 \left(\pi_0 + \pi_1 + \pi_2\right)$$

subject to:

$$\pi_0 + q_0 i_0 = n_0 + E\left(b_1\right) i_0, \tag{7}$$

$$\pi_{1s} + q_{1s} k_s = q_{1s} i_0 + x_s i_0 - b_{1s} i_0 + b_{2s} k_s \quad \forall s\text{: partial or no liquidation,} \tag{8}$$

$$\pi_{1s} = q_{1s} i_0 - b_{1s} i_0 \quad \forall s\text{: total liquidation in period 1,} \tag{8L}$$

$$\pi_{2s} = A k_s - b_{2s} k_s \quad \forall s\text{: partial or no liquidation,} \tag{9}$$

$$\pi_{2s} = 0 \quad \forall s\text{: total liquidation in period 1,} \tag{9L}$$

$$0 \leq b_{1s} \leq \theta q_{1s} \qquad \forall s, \tag{10}$$

$$0 \leq b_{2s} \leq \theta q_{2s} \qquad \forall s. \tag{11}$$

Equation (7) represents the intermediary's period 0 budget constraint: investment costs and any profits taken by the intermediary in period 0 must be financed by its endowment (initial net worth) and borrowing from consumers.[11] In period 1, provided that the investment project is continued (i.e. provided that the intermediary does not go into total liquidation), the intermediary's budget constraint is given by (8): financing is provided by start of period assets at their market value ($q_{1s}i_0$) and net period 1 borrowing ($b_{2s}k_s - b_{1s}i_0$), adjusted for the revenue surplus or shortfall, $x_s i_0$. Period 2 profits in this case are then given by (9). By contrast, if the intermediary goes into total liquidation in period 1, it sells all of its capital at the market price, yielding $q_{1s}i_0$ in revenue. Therefore, its period 1 profits are given by (8L), while period 2 profits are zero (equation (9L)). Finally, note that (10) and (11) simply represent combined and simplified versions of the borrowing constraints, (3)-(6).

This optimisation problem can immediately be simplified. Since expected returns on investment are always high, it is clear that the intermediary will never take any profits until period 2 unless it goes into total liquidation.[12] Therefore $\pi_0 = 0$ in (7) and $\pi_{1s} = 0$ for all s in (8). Moreover, given that the high return between periods 1 and 2 is certain, intermediaries wish to borrow as much as possible in period 1. So (11) binds at its upper bound and $b_{2s} = \theta q_{2s}$. Finally, the asset price is only endogenous in period 1: $q_0 = 1$ because of the large supply of consumption goods in period 0 and we set $q_{2s} = 1$ for all

[11] Both this and the other budget constraints must bind by local non-satiation.

[12] Period 1 profits may be positive if the intermediary goes into total liquidation because it does not need to pay xi_0 if it shuts down and can retain any proceeds remaining from asset sales after outstanding liabilities have been paid. Note that total profits are still increasing in x; the only difference is that if the intermediary continues to operate, it takes its (higher) profits in period 2 and nothing in period 1.

s.[13] Therefore, we can rewrite the intermediary's optimisation problem as:

$$\max_{i_0,\{k_s\},\{b_{1s}\}} E_0\left(\pi_1 + \pi_2\right)$$

subject to:

$$i_0 = n_0 + E\left(b_1\right) i_0, \tag{12}$$

$$q_{1s}k_s = q_{1s}i_0 + x_s i_0 - b_{1s}i_0 + \theta k_s \qquad \forall s\text{: partial or no liquidation,} \tag{13}$$

$$\pi_{1s} = q_{1s}i_0 - b_{1s}i_0 \qquad \forall s\text{: total liquidation in period 1,} \tag{8L}$$

$$\pi_{2s} = Ak_s - \theta k_s \qquad \forall s\text{: partial or no liquidation,} \tag{14}$$

$$\pi_{2s} = 0 \qquad \forall s\text{: total liquidation in period 1,} \tag{9L}$$

$$0 \le b_{1s} \le \theta q_{1s} \qquad \forall s. \tag{10}$$

2.2 *Multiple Equilibria and Systemic Crises: Intuition*

Before solving the intermediary's optimisation problem, we graphically illustrate how multiple equilibria and systemic financial crises arise in the model. Faced with a negative realisation of x, intermediaries may be forced to sell a portion of their capital to the traditional sector in period 1 to remain solvent. In these fire sale states, $i_{1s} = 0$ and, using (1), $k_s = i_0 - k_s^S = i_0 - k_s^T$, where $k_s^S = k_s^T \le i_0$. Provided that intermediaries remain solvent, we can substitute this expression into (13) and rearrange to obtain the inverse supply function for capital in the traditional sector:

$$q_{1s} = \frac{\left(b_{1s} - x_s - \theta\right) i_0}{k_s^T} + \theta. \tag{15}$$

From (15), it is clear that the supply function is downward sloping and convex. The intuition for this is that when the asset price falls, intermediaries are forced to sell more

[13] We set $q_{2s} = 1$ because we wish to allow for borrowing between periods 1 and 2 without setting up an infinite horizon model. This assumption can be justified by assuming that period 2 returns are realised in two stages. In the first stage, the intermediaries must control the capital and $(A-1)k_s$ units of the consumption good are realised; in the second stage, k_s units are realised irrespective of who controls the capital. Between these stages, intermediaries must repay consumers with consumption goods and, if necessary, a portion of their capital – if they do not, their capital will be seized. Since everyone can gain a return from capital at this point, its marginal value is one, and hence $q_{2s} = 1$.

capital to the traditional sector to remain solvent; the more the asset price falls, the more capital needs to be sold to raise a given amount of liquidity. Equation (15) holds for all $k_s^T < i_0$. But if intermediaries sell all of their capital and go into liquidation, the supply of capital to the traditional sector is simply given by:

$$\left(k_s^T\right)^L = i_0. \tag{16}$$

Meanwhile, since the traditional sector is perfectly competitive, the inverse demand function for capital sold by intermediaries follows directly from (2):

$$q = F'\left(k^T\right) = 1 - 2\alpha k^T. \tag{17}$$

This function is downward sloping and linear due to linearly decreasing returns to scale in the traditional sector. Combining (15), (16) and (17) yields the equilibrium asset price(s) in fire sale states.

The supply and demand functions are sketched in $\left(q, k^T\right)$ space in Figure 2. As can be seen, there is the potential for multiple equilibria in fire sale states. In particular, if the supply schedule is given by S'', there are three equilibria: R$''$, U and C. From (15), $S(0) > 1$ for all supply schedules. Therefore, U is unstable but the other two equilibria are stable. Point C corresponds to a crisis: intermediaries go into liquidation, firms shut down, and all capital is sold to the traditional sector, causing the asset price to fall substantially. By contrast, at R$''$, fire sales are limited and the asset price only falls slightly – we view this as a 'recession' equilibrium since intermediaries remain solvent and firms continue to operate in the productive sector.

The actual outcome between R$''$ and C is determined solely by beliefs: if intermediaries believe *ex ante* (before the realisation of the shock) that there will be a systemic crisis in states for which there are multiple equilibria, a crisis will indeed ensue in those states; if they believe *ex ante* that there will only be a 'recession' in those states, then that will be the outcome. Moreover, their *ex ante* investment and borrowing decisions depend on their beliefs. Therefore, multiple equilibria arise *ex ante*: after beliefs have

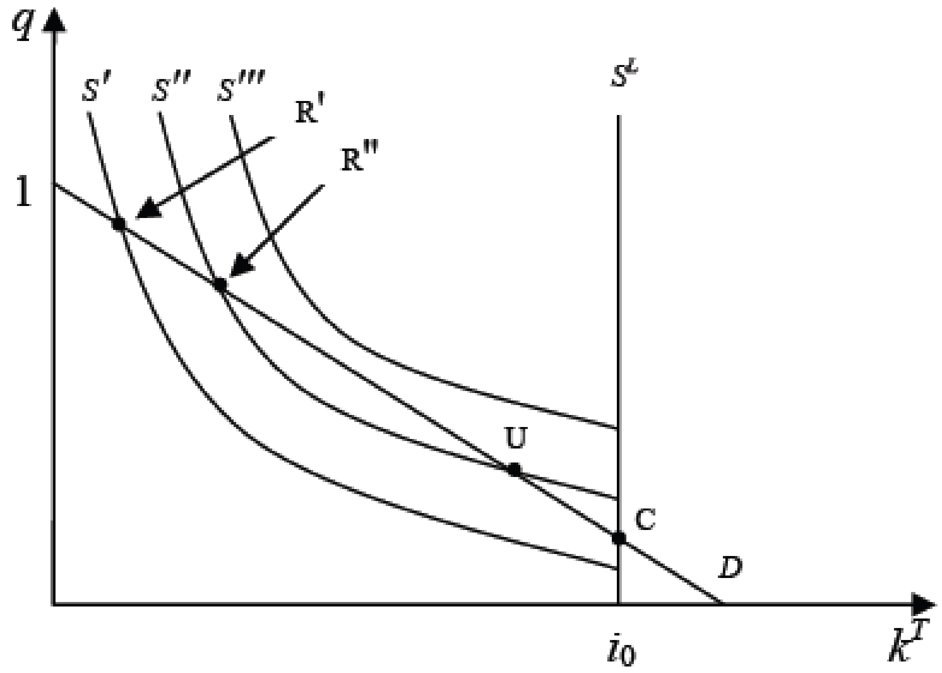


Figure 2: Demand and Supply for Capital in the Traditional Sector

been specified (at the start of period 0), investment and borrowing decisions will be made contingent on those beliefs and the period 1 equilibrium will be fully determinate, even in states for which there could have been another equilibrium.

However, multiple equilibria and systemic crises are not always possible in fire sale states. Specifically, if the supply schedule is given by S', R$'$ is the unique equilibrium and there can never be a systemic crisis, regardless of beliefs. From (15), it is intuitively clear that this is more likely to be the case when the negative x shock is relatively mild. By contrast, if the shock is extremely severe, a crisis could be inevitable – supply schedule S''' depicts this possibility.

2.3 *The Competitive Equilibrium*

We now proceed to solve the model for both 'optimistic' and 'pessimistic' beliefs. Suppose that all agents form a common exogenous belief at the start of period 0 about what equilibrium will arise when multiple equilibria are possible in period 1: if beliefs are 'optimistic', agents assume that there will not be a crisis unless it is inevitable (i.e. unless the supply schedule resembles S'''); if beliefs are 'pessimistic', agents assume that

State	Realisation of x_s	Description of Outcome
'Good'	$x_s > 0$	Intermediaries do not sell any capital. There is no production in the traditional sector.
'Recession'	x^C or $x^M \leq x_s \leq 0$	Intermediaries sell a portion of their capital but remain solvent (i.e. there are only limited fire sales). Firms continue to operate in the productive sector, but with a lower capital stock than in 'good' states. There is some production in the traditional sector.
'Crisis'	$x_s < x^C$ or x^M	Intermediaries sell all of their capital and go into liquidation. Firms operating in the productive sector shut down. Production only takes place in the traditional sector.

Table 1: Summary of Outcomes

if there is a possibility of a crisis, it will indeed happen. Then, as shown in Appendix A, the competitive equilibrium is characterised by the following repayment ratios associated with each possible state, x_s, where the precise thresholds ($\widehat{x}$, $\widehat{x} - \theta\widehat{q}$ and x^C) depend on beliefs and the distribution of shocks:

$$\text{if } \widehat{x} < x_s \text{, then } b_{1s} = \theta q_{1s}, \tag{18}$$

$$\text{if } \widehat{x} - \theta\widehat{q} < x_s < \widehat{x} \text{, then } b_{1s} = \theta\widehat{q} - (\widehat{x} - x_s), \tag{19}$$

$$\text{if } x^C < x_s < \widehat{x} - \theta\widehat{q} \text{, then } b_{1s} = 0, \tag{20}$$

$$\text{if } x_s < x^C \text{, then } b_{1s} = \theta q^C = \max[\theta(1 - 2\alpha i_0), 0]. \tag{21}$$

Expressions (18)-(20) correspond to similar expressions in Lorenzoni (2005), though the actual thresholds differ. However, (21) is specific to our model and reflects the possibility of systemic financial crises in our setup.

Apart from noting that $\widehat{x} \leq 0$ (since intermediaries will never choose to borrow less than the maximum against states where the realised x is positive), relatively little can be said about the precise location of the thresholds without specifying how the shock is distributed. Section 3 determines these thresholds, initial investment, and the state-contingent asset price for a specific distribution.

2.4 *Discussion of the Competitive Equilibrium*

Since expected future returns are positive, the competitive equilibrium always exhibits a high level of credit-financed investment in period 0. As summarised in Table 1, subsequent outcomes depend on the realisation of x. In 'good' states, x is positive, investment and credit growth remain strong in period 1, and the economy benefits from high returns in period 2. Of more interest for our analysis are the 'recession' and 'crisis' states in which x is negative. To further clarify what happens in these cases, we sketch the period 1 repayment ratio, b_1, and asset price, q_1, against x in Figures 3 and 4 respectively. For illustrative purposes, we present the cases of 'optimistic' and 'pessimistic' beliefs on the same diagram, adding an additional threshold, x^M, to reflect the range of x for which multiple equilibria are possible.[14] However, it is important to bear in mind that the thresholds themselves are endogenous to beliefs.

To explain the repayment ratio function in Figure 3, consider what happens when there is a negative x shock (for positive x, $q_1 = 1$, implying that $b_1 = \theta$). As noted above, if the intermediary goes into liquidation as a result of the shock (i.e., if $x_s < x^C$ or x^M, depending on beliefs), it does does not need to pay the cost xi_0. In this case, it sells all of its capital at the prevailing market value and repays this 'scrap value' to consumers. Although it may seem unusual that repayments are positive in 'crisis' states (and potentially higher than in 'recession' states), this is entirely optimal. Intuitively, intermediaries have no need for liquidity in 'crisis' states because they shut down and do not pay the cost xi_0. By increasing repayments to consumers in these states, they are able to increase their period 0 borrowing. Since period 0 investment is expected to be profitable, it is, therefore, optimal for intermediaries to promise to repay the entire 'scrap value' of the project to consumers in 'crisis' states.

If, however, the intermediary wants to avoid total liquidation following a negative shock, it must find a way of financing the cost xi_0. Given that it always chooses to borrow the maximum amount it can between periods 1 and 2, the cost can be financed either by reducing repayments to consumers in adverse states or by selling a portion of

[14] As for the other thresholds, the location of x^M cannot be computed without specifying the distribution of the shock. However, Figure 2 and the associated discussion clearly illustrate how multiple equilibria are only possible over a certain range of x.

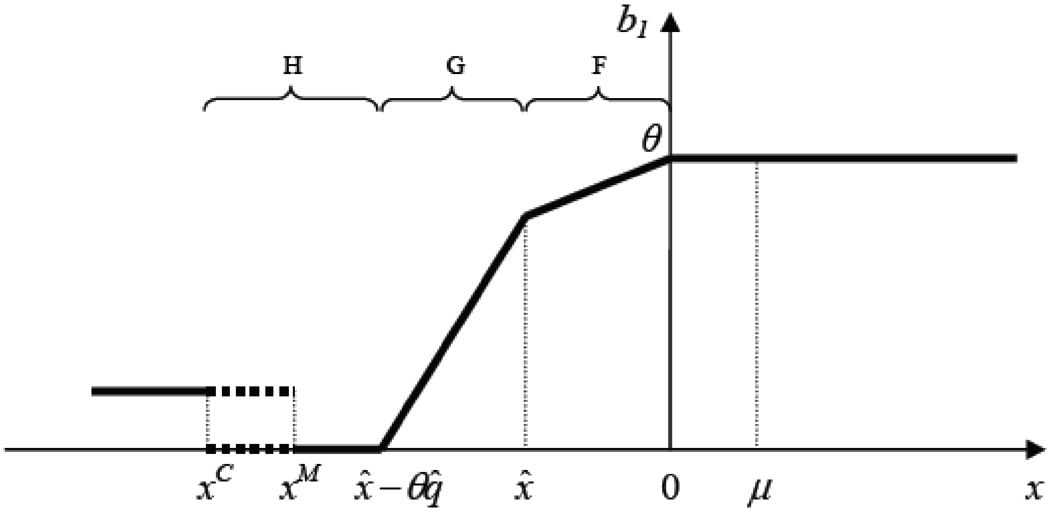


Figure 3: The Repayment Ratio as a Function of the Shock

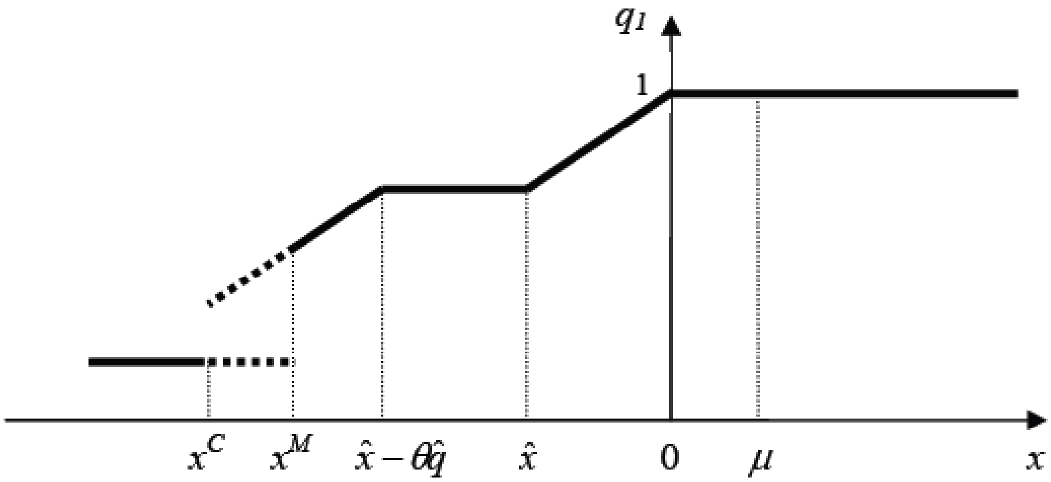


Figure 4: The Asset Price as a Function of the Shock

its capital.

The first option reduces expected repayments to consumers (i.e. $E(b_1)$), lowering the amount that the intermediary can borrow in period 0 (see equation (12)) and therefore reducing returns in 'good' states. The expected cost associated with doing this is constant. By contrast, the cost of the second option increases as the asset price falls. So, for mild negative shocks in region F of Figure 3, it is better to sell capital because the asset price remains relatively high. The borrowing / repayment ratio in these states remains at its maximum, but this maximum falls slowly as the asset price falls (see equations (5) and (18)).

However, when shocks are more severe and fall in region G, the costs of selling capital are so high that it becomes better to reduce repayments to consumers than to sell further capital – this is reflected in (19). Eventually, however, the scope for reducing repayments is fully exhausted and the only way to finance the cost is to sell further capital even though the asset price is relatively low (region H). It is at this point that the $b_{1s} > 0$ constraint bites: intermediaries would ideally like to receive payments from consumers in these extremely bad states but are prevented from doing so by limited commitment on the consumer side.[15]

Since the asset price, q_1, only changes when the amount of capital being sold changes, the intuition behind Figure 4 follows immediately. For positive x, no capital is ever sold, so the asset price remains at one. However, for negative (but non-crisis) values of x, the asset price falls over those ranges for which intermediaries finance xi_0 by selling additional capital (i.e. for $\widehat{x} < x_s < 0$ and $x_s < \widehat{x} - \theta\widehat{q}$). Meanwhile, in crises, intermediaries sell all of their capital and the asset price is determined by substituting (16) into (17), which gives $q^C = 1 - 2\alpha i_0$. If this expression is negative, returns to capital in the traditional sector fall to zero before all the available capital is being used. In this case, the leftover capital has no productive use in the economy, and $q^C = 0$.

2.5 *The Constrained Efficient Equilibrium, Efficiency, and the Source of the Externality*

We can show that the competitive equilibrium is constrained inefficient by solving the problem faced by a social planner who maximises the same objective function as intermediaries and is subject to the same constraints, but who does not take prices as given. Under certain mild conditions (see Appendix B), the social planner can obtain a welfare improving allocation by reducing intermediaries' borrowing and investment. More specifically, the social planner implements a reduction in borrowing against certain states that has no direct effect on intermediaries' welfare. But it has a potentially

[15] Since early investment is expected to be profitable, intermediaries have no incentive to set aside liquid resources in period 0 to self-insure against extremely bad states in period 1. But even if some self-insurance were optimal, asset sales would still be forced for sufficiently severe shocks.

important indirect effect: by reducing investment, the amount of capital that has to be sold in fire sale states is reduced, and this both reduces the negative effects of asset price falls, and lowers the likelihood and severity of crises.

The competitive equilibrium thus exhibits over-borrowing and over-investment relative to the constrained efficient equilibrium. In particular, if we view the situation with no frictions (i.e. without borrowing constraints (3)-(6)) as corresponding to the first-best outcome and the constrained efficient equilibrium as the second-best, then the competitive allocation is fourth-best. This is because policy intervention could feasibly achieve a third-best outcome even if the second-best allocation cannot be attained.

As noted earlier, the limited commitment and potential default to which financial contracts are subject is the key friction in this model. It is straightforward to show that the critical constraint is (3): if this were relaxed, the competitive equilibrium would be efficient and there would never be systemic crises because intermediaries would be able to obtain additional payments from consumers in times of severe stress (i.e. when $x_s < \widehat{x} - \theta\widehat{q}$) rather than being forced to sell capital. However, when coupled with decreasing returns to capital in the traditional sector, the presence of this constraint introduces an asset fire sale externality: intermediaries do not internalise the negative effects on asset prices that their own fire sales have. By tightening their budget constraints further, these asset price falls force other intermediaries to sell more capital than they would otherwise have to. In extreme cases, this externality is the source of systemic crises.

3 Comparative Statics

We now analyse the effects of financial innovation and changes in macroeconomic volatility on the likelihood and potential scale of systemic crises. This necessitates an assumption about beliefs so that the cut-off value of x below which crises occur is determinate. Accordingly, we suppose that agents have 'optimistic' beliefs, so that crises only occur when they are inevitable.[16]

The shock x is assumed to be normally distributed with mean μ and variance σ^2,

[16] All of our qualitative results continue to hold if agents have 'pessimistic' beliefs.

where $\mu > 0$. Since analytical solutions for thresholds are unavailable, we present the results of numerical simulations. In our baseline analysis, we assume the following parameter values: $A = 1.5$; $n_0 = 1$; $\mu = 0.5$; $\sigma = 0.5$; $\theta = 0.75$; $\alpha = 0.05$. We then consider the effects of varying σ, θ and α. The empirical relevance of the parameters used is discussed in section 3.3.[17]

We measure the likelihood of a crisis by $H(x^C) = \Pr[x < x^C]$ and its scale (impact) in terms of the asset price, q^C, which prevails in it.[18] Lower values of q^C correspond to more serious crises. To motivate q^C as a measure of the impact of crises, recall that in period 0, consumption goods are turned into capital goods one for one. If some capital goods end up being used in the less-productive sector to produce consumption goods (as happens in a crisis), fewer consumption goods can be produced than were used to buy those capital goods initially. Since a lower q corresponds to reduced returns on the marginal unit of capital in the traditional sector and hence less production of the consumption good from the marginal capital good, the loss associated with a crisis increases as q^C falls. Moreover, lower values of q^C correspond to greater asset price volatility in the economy, further suggesting that it may be an appropriate measure of the scale of systemic instability.

3.1 *Changes in Macroeconomic Volatility*

We interpret a change in macroeconomic volatility as affecting σ. Since x is linked to revenue shortfalls and surpluses, it is reasonable to assume that a reduction in output and inflation volatility (as is likely to be associated with a general reduction in macroeconomic volatility) corresponds to a fall in the standard deviation of x.

Intuitively, a reduction in σ will lower the probability of crises since extreme states become less likely. This is borne out in Figure 5(a). However, provided that the mean, μ, is sufficiently above zero and the variance is not too large, a lower standard deviation also makes 'recession' states less likely to occur. As a result, expected repayments to

[17] The *Matlab* code used for the simulations is available on request from the authors. Robustness checks were also performed by varying the parameters over a range of values.

[18] Recall that crises are associated with total liquidation. So, although the distribution of shocks, $H(x)$, is continuous, there is only one crisis price, q^C, for all x less than x^C.

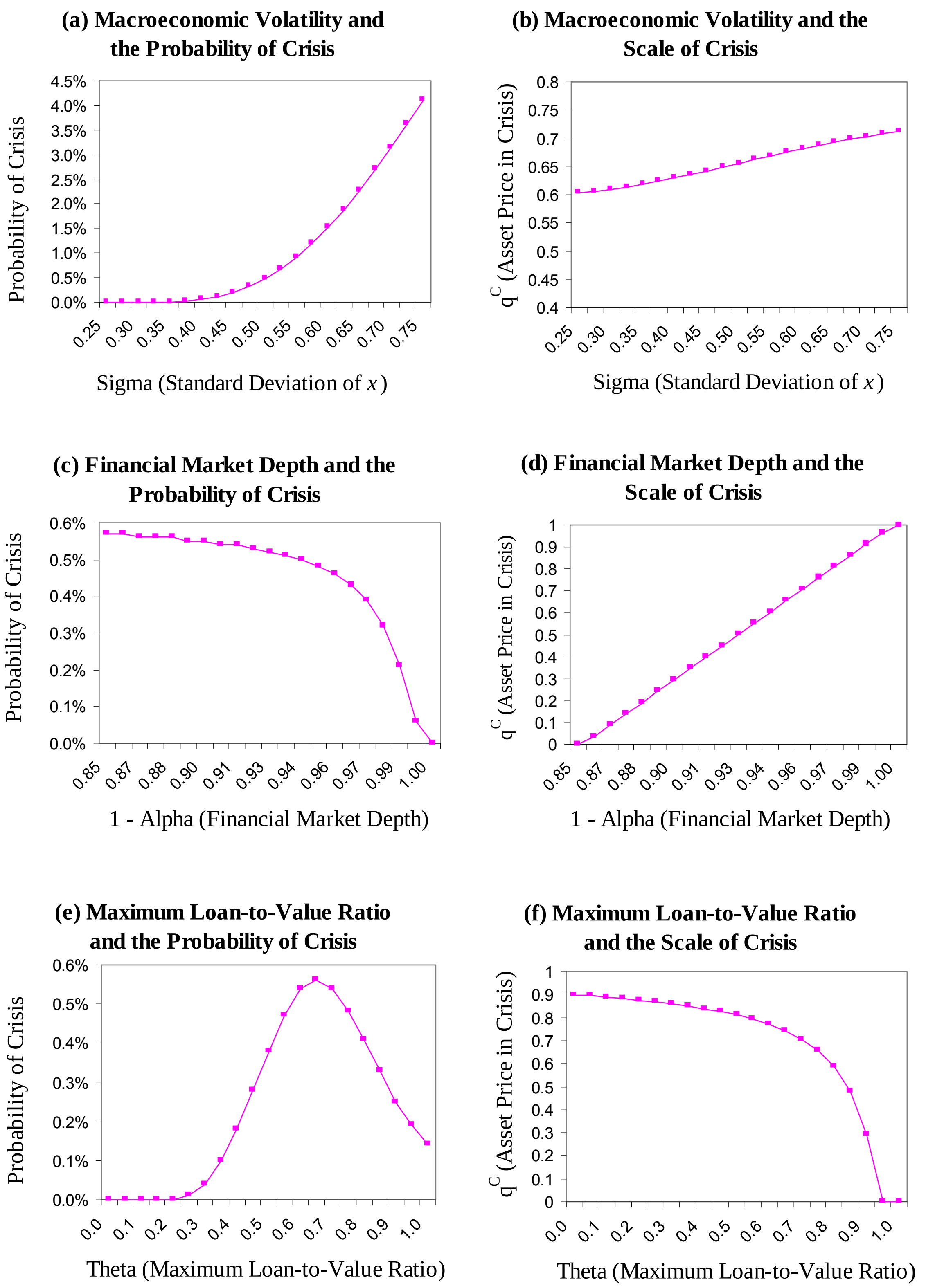


Figure 5: Comparative Static Results

consumers, $E(b_1)$, are higher, meaning that intermediaries can borrow more in period 0. Therefore, initial investment, i_0, is higher. But this means that if a crisis then does arise, more capital will be sold to the traditional sector, the asset price will be driven down further, and the crisis will have a greater impact. This is shown in Figure 5(b) and can also be seen by considering a rightward shift of S^L in Figure 2.[19]

3.2 *The Impact of Financial Innovation*

We have already argued that financial innovation and recent developments in financial markets can be interpreted as implying higher maximum loan-to-value ratios (higher values of θ) and greater financial market depth (lower values of α). Assuming that the initial value of θ is not particularly low, Figure 6(a) illustrates how these changes have made crises less likely (darker areas in the chart correspond to a higher crisis frequency). But from Figure 6(b), it is apparent that the severity of crises may have increased (darker areas correspond to a more severe crisis).

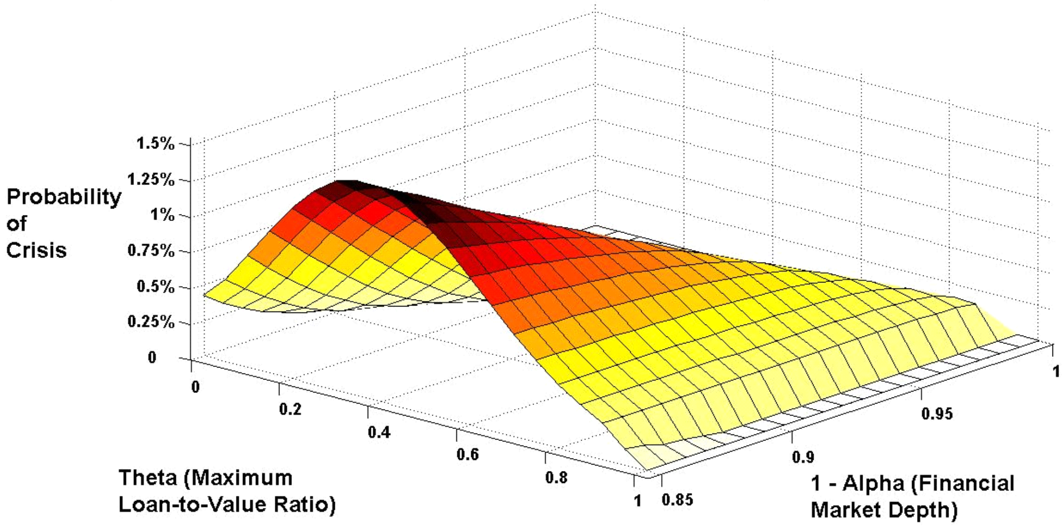
To understand the intuition behind these results, we isolate the individual effects of changes in α and θ. Figures 5(c) and 5(d) suggest that a reduction in α reduces both the likelihood and scale of crises. This is intuitive. If the secondary market for capital is deeper, shocks can be better absorbed and, in the context of Figure 2, the demand curve in the traditional sector is flatter. As a result, crises are both less likely and less severe.[20]

By contrast, Figures 5(e) and 5(f) suggest that an increase in θ increases the severity of crises and has an ambiguous effect on their probability. This is demonstrated more formally in Appendix C. Intuitively, a rise in θ enables intermediaries to borrow more. Therefore, i_0 is higher, and crises will be more severe if they occur. Greater borrowing in period 0 clearly serves to increase the probability of crises as well. However, a rise in θ also means that intermediaries have greater access to liquidity in period 1: specifically,

[19]If μ is very close to zero and/or σ is very large, it is possible for a reduction in σ to make 'recession' states more likely. This can potentially lead to a reduction in $E(b_1)$ and hence i_0, thus reducing the impact of crises. Since the numerical results suggest that this only happens for fairly extreme combinations of the mean and variance, we view the case discussed in the main text as being more likely. However, this feature does have the interesting implication that crises could be most severe in fairly stable and extremely volatile economies.

[20]This analysis assumes that secondary markets continue to function with the onset of a crisis. However, α itself could be endogenous and change during periods of stress. So reductions in α in benign times may have little effect on the severity of crises.

(a) Maximum Loan-to-Value Ratio, Financial Market Depth and the Probability of Crisis



(b) Maximum Loan-to-Value Ratio, Financial Market Depth and the Scale of Crisis

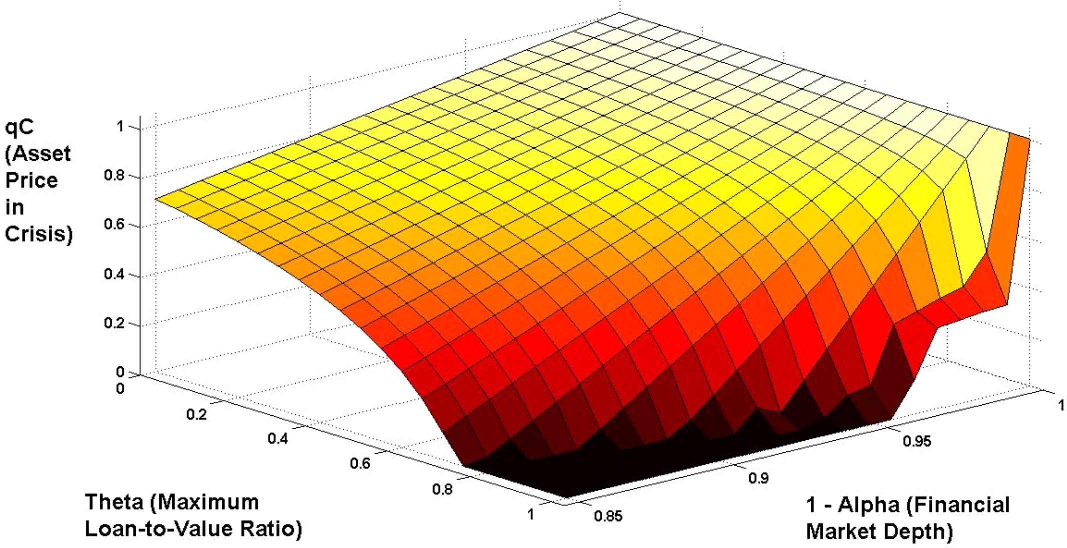


Figure 6: Financial Innovation and the Probability and Scale of Crises: 3D Charts

they have more scope to reduce period 1 repayments to consumers. This effect means that they are less likely to go into total liquidation, making crises less likely.

Figure 5(e) shows that crises are most frequent for intermediate values of θ, suggesting that middle-income emerging market economies may be most vulnerable to systemic instability.[21] By contrast, countries with extremely well-developed or very underdeveloped financial sectors, with high / low maximum loan-to-value ratios, are probably less vulnerable to crises.

3.3 *Comments on the Quantitative Results*

Although our numerical analysis is intended to be illustrative, the baseline case is broadly consistent with several features of the data. As would be expected, the leverage ratio of assets to equity implied by the model is closely tied to the value of θ. With θ set to be 0.75, the implied leverage ratio is 3.5, which is reasonably close to the estimate of 4.9 for average hedge fund leverage over 1996-2004 reported by McGuire *et al.* (2005).

The mean and variance of the shock are chosen in relation to each other and are key determinants of the likelihood of 'recessions' and crises. If a period is taken as one year, the baseline parameter values yield 'recessions' once every six and a half years and crises once every 200 years. In 'recession' states, the average short-run loss which intermediaries have to finance is 24% of the initial amount invested; in crises, the reinvestment cost needed to continue operations (which intermediaries choose not to pay) is almost as much as the initial amount invested. Price falls in adverse states are strongly influenced by α – in the baseline calibration, the average price discount in 'recession' states is 17%, while the price falls by 35% in crises. These figures are broadly consistent with the 30% price discount identified by Pulvino (1998) for commercial aircraft sales in depressed markets and the 7.9% price discount for fire sale stocks reported by Coval and Stafford (2007), especially when we consider that equities are amongst the most liquid assets, whilst aircraft are probably amongst the most illiquid assets.

[21] Aghion *et al.* (2004) present a similar result but their approach is quite different, focussing on the effects of fluctuating real exchange rates and international capital flows in a small open economy model.

3.4 *Discussion*

The comparative static analysis highlights the potential risk of more severe crises as a result of financial innovation and greater macroeconomic stability. But this should not necessarily be taken to imply that these developments are undesirable. In particular, higher values of θ and lower values of σ both imply greater investment in period 0 and, as such, may increase welfare.

All of our results were obtained under the assumption that θ is not state-contingent. But *ex post* changes in θ in period 1 can affect outcomes. In particular, it is clear from (15) that when i_0 is strictly greater than k_s^T, an unanticipated increase in θ in period 1 states with a negative x will shift the supply curve for capital in the traditional sector to the left. As a result, there will be fewer cases in which crises are inevitable. In addition, the price fall in 'recession' states will be lower. Intuitively, the *ex post* increase in θ enables intermediaries to access more liquidity in period 1, meaning that they do not need to sell as many assets to the traditional sector to continue operations. On the other hand, falling maximum loan-to-value ratios during downturns could have detrimental effects.

This result suggests that a rule to increase θ in adverse states may be welfare-improving, though a full analysis would clearly require solving the model under the assumption that, when making initial investment decisions, intermediaries know that θ may be adjusted in period 1. As such, the model illustrates how there may sometimes be scope for policymakers to promote liquidity. One specific approach, discussed by Borio *et al.* (2001), is the pursuit of discretionary policy towards collateral valuation practices during periods of stress. For example, as noted by Borio (2004), supervisory authorities in Japan lowered margin requirements and relaxed lending limits on collateral assets in order to alleviate liquidity constraints and contain distress selling during the 1987 stock market crash. More generally, the welfare consequences of policies that induce market participants to hold liquidity cushions at business-cycle frequencies – building up liquidity during booms and drawing it down during recessions – merit closer investigation.

4 Conclusion

This paper analysed a theoretical general equilibrium model of intermediation with financial constraints and state-contingent contracts containing a clearly defined pecuniary externality associated with asset fire sales during periods of stress. After showing that this externality was capable of generating multiple equilibria and systemic financial crises, we considered the effects of changes in macroeconomic volatility and developments in financial markets on the likelihood and severity of crises. Together, our results suggest how greater macroeconomic stability and financial innovation may have reduced the probability of systemic financial crises in developed countries in recent years. But these developments could have a dark side: should a crisis occur, its impact could be greater than was previously the case.

The paper sheds light on cross-country variation in the likelihood and scale of financial crises. Macroeconomic volatility is generally higher in developing countries than in advanced economies but maximum loan-to-value ratios are invariably lower. Given this, our results predict that crises in emerging market economies should be more frequent but less severe than in developed countries. The first of these assertions is clearly borne out by the data (Caprio and Klingebiel, 1996, Table 1; Demirguc-Kunt and Detragiache, 2005, Table 2). Although the second is more difficult to judge given the rarity of financial crises in developed countries in recent years, the length and depth of the Japanese financial crisis of the 1990s suggests that such intuition is plausible. Moreover, in terms of output losses, Hoggarth *et al.* (2002) find that crises in developed countries do indeed tend to be more costly than those in emerging market economies.

Appendix A: The Competitive Equilibrium

In this appendix, we solve the model for the competitive equilibrium when all agents have ‘optimistic’ beliefs about what equilibrium will arise in states in which multiple equilibria are possible. Specifically, they believe that crises only happen when they are inevitable and never occur when there are multiple equilibria. If agents have ‘pessimistic' beliefs, the derivation proceeds along very similar lines.

Conditional on beliefs, the equilibrium is unique, and can be fully characterised by the three cut-off values for the aggregate shock x shown in expressions (18)-(21). These cut-offs determine four intervals in the distribution of x (i.e. in the distribution of possible states). In each of these intervals, intermediaries' incentives to protect their net worth, and hence their decisions about optimal repayments, will be different. We show how the equilibrium can be fully characterised by these three cut-off points and how, conditional on beliefs, it is unique.

Define the subset $\mathbf{C}$ as the (endogenous) set of states where there is a crisis. Then the return, z_s, that intermediaries obtain in period 2 in state s from one unit of their net worth in state s in period 1 is given by:

$$z_s = \begin{cases} \frac{A-\theta}{q_{1s}-\theta} & \forall s \notin \mathbf{C} \\ 1 & \forall s \in \mathbf{C} \end{cases}. \tag{22}$$

To derive this expression, note that in non-crisis states in period 1, a given amount of net worth, n_1, can be leveraged to obtain a total investment by intermediaries of $q_{1s}k_s = n_1 + \theta k_s$. In other words, each unit of net worth is leveraged by a factor of $1/\left(q_{1s} - \theta\right)$. Since the return per unit of capital after payment of liabilities is $A - \theta$ (recall that $b_{2s} = \theta$), return per unit of net worth in non-crisis states is therefore $\left(A - \theta\right)/\left(q_{1s} - \theta\right)$. By contrast, in crisis states, intermediaries do not invest, so the marginal return to net worth is just its consumption value of one.

Meanwhile, the return, z_0, that intermediaries obtain in period 2 by investing one unit of their net worth in period 0 is given by:

$$z_0 = E_{s\notin\mathbf{C}}\left[z\frac{x+q-b_1}{1-E(b_1)}\right]\Pr[s\notin\mathbf{C}] + E_{s\in\mathbf{C}}\left[\frac{q-b_1}{1-E(b_1)}\right]\Pr[s\in\mathbf{C}]. \tag{23}$$

This is the expected value of the product of period 1 and period 2 returns. The period 1 return may be explained along similar lines to the period 2 return. The factor by which intermediaries leverage one unit of period 0 net worth to purchase capital is $1 - E\left(b_1\right)$. In non-crisis states, the return per unit of capital is $x_s + q_{1s} - b_{1s}$. However, since intermediaries that fully liquidate do not pay the cost $x_s i_0$, the return per unit of capital in crisis states is $q_{1s} - b_{1s}$.

States can be divided into four sets: $S_1 = \{s : 1 < z_s < z_0\}$, $S_2 = \{s : z_s = z_0\}$, $S_3 = \{s : z_s > z_0\}$, and $\mathbf{C} = \{s : z_s = 1 < z_0\}$.We want to show that these sets cover the whole distribution of x, with S_1 covering states from $+\infty$ to $\widehat{x}(< 0)$, S_2 from $\widehat{x}$ to $\widehat{x} - \theta\widehat{q}$, S_3 from $\widehat{x} - \theta\widehat{q}$ to x^C, and $\mathbf{C}$ from x^C to $-\infty$.

Consider a state s that belongs to S_1. We want to show that if $x_{s'} > x_s$, then $s' \in S_1$. In state $s \in S_1$, borrowing will be at its maximum possible level in period 0 ($b_{1s} = \theta q_{1s}$) because $z_0 > z_s$, and the price of capital will satisfy $q_{1s} = F'[\max(k_s^T, 0)]$. If $x_{s'} > x_s > 0$, then there are no fire sales and $q_{1s} = q_{1s'} = 1$, and $z_s = z_{s'}$. If $0 > x_{s'} > x_s$, then $k_{s'}^T < k_s^T$, $q_{1s} < q_{1s'}$ and $z_{s'} < z_s$. In both cases, $z_{s'} < z_0$ and hence s' belongs to S_1.

The threshold for x that separates S_1 and S_2 is $\widehat{x}$. It is the value for which, in equilibrium, $z_0 = z_s$ and there is maximum borrowing ($q_{1s} = \widehat{q}$ is the equilibrium price in that state). For all states in $S_2 = \{s : z_s = z_0\}$, q_{1s} has to be constant, and given that i_0 is constant in all states in S_2, the amount borrowed in each state is pinned down and given by $b_{1s} = \theta\widehat{q} - (\widehat{x} - x_s)$. The second cut-off, $\widehat{x} - \theta\widehat{q}$, is the value of x for which $b_{1s} = 0$ and $z_s = z_0$. As x decreases beyond $\widehat{x} - \theta\widehat{q}$, the repayment / borrowing ratio cannot be reduced any further. Therefore, more capital is sold in the secondary market, implying that $q_{1s} < \widehat{q}$ and hence $z_s > z_0$. Following the same logic as when we show that all values above $\widehat{x}$ belong to S_1, it is straightforward to show that all values below $\widehat{x} - \theta\widehat{q}$ but above the crisis threshold, x^C, belong to S_3. (It is important to note at this point that we are assuming that whenever it is possible to have multiple equilibria, 'optimistic' self-fulfilling beliefs imply that the 'recession' equilibrium arises rather than the 'crisis' equilibrium. We do not specify the precise set of multiple equilibria states, as this set is itself endogenous and a function of beliefs.)

To complete the characterisation, we need to show that there is a threshold, x^C, below which crises are unavoidable, and find conditions under which this threshold is lower than $\widehat{x} - \theta\widehat{q}$. The solution for x^C is obtained by solving the system of two equations that results from equating the demand and supply curves and their slopes. It is given by:

$$x^C = -\left[\frac{(1-\theta)^2}{8\alpha i_0} + \theta\right]. \tag{24}$$

An exact analytical condition for x^C to be lower than $\widehat{x} - \theta\widehat{q}$ requires an assumption about the distribution of x. In our numerical exercises we check that this condition is satisfied, finding that it is for most parameter values.

Appendix B: The Social Planner's Solution

The social planner's optimisation problem is given by:

$$\max_{i_0,\{k_s\},\{b_{1s}\}} E_0\left(\pi_1 + \pi_2\right) = \max_{i_0,\{k_s\},\{b_{1s}\}} E_{s\notin\mathbf{C}}\left[\frac{A-\theta}{q-\theta}(x+q-b_1)i_0\right]\Pr[s \notin \mathbf{C}]$$
$$+ E_{s\in\mathbf{C}}\left[\left(q-b_1\right)i_0\right]\Pr[s \in \mathbf{C}]$$

subject to:

$$i_0 = n_0 + E\left(b_1\right)i_0 - \tau, \tag{25}$$

$$k_s^T q_{1s} = -\left(x_s - b_{1s}\right)i_0 - \left(i_0 - k_s^T\right)\theta \quad \forall s\text{: partial or no liquidation } (s \notin \mathbf{C}), \tag{26}$$

$$0 \leq b_{1s} \leq \theta q_{1s} \quad \forall s, \tag{10}$$

and:

$$E\left[3e + \tau + F(k^T) - qk^T - w\right] \geq U^{CE}, \tag{27}$$

where $\mathbf{C}$ is the set of crisis states, U^{CE} is the utility of consumers under the competitive equilibrium, τ is a transfer from intermediaries to consumers, $F\left(k^T\right) - qk^T$ represents profits to consumers from production in the traditional sector, $w = -\lambda x$ is the cost of a financial crisis to consumers, and $0 < \lambda < 1$.

Condition (27) requires that consumers are at least as well off in the constrained efficient equilibrium as in the competitive equilibrium. To satisfy this condition, the social planner implements any necessary transfer, τ, from intermediaries to consumers in period 0. The key difference between the social planner and representative intermediary problems is that the social planner does not take the asset price, q_{1s}, as given.

Since $q_{1s} = F'\left(k_s^T\right)$ and since $k^T = i_0$ in crisis states, the social planner's problem can be rewritten as:

$$\max_{i_0,\{k_s\},\{b_{1s}\}} E_0\left(\pi_1+\pi_2\right) = \max_{i_0,\{k_s\},\{b_{1s}\}} E_{s\notin\mathbf{C}}\left\{\frac{A-\theta}{F'\left(k^T\right)-\theta}\left[x+F'\left(k^T\right)-b_1\right]i_0\right\}\Pr[s\notin\mathbf{C}]$$
$$+E_{s\in\mathbf{C}}\left\{\left[F'\left(i_0\right)-b_1\right]i_0\right\}\Pr[s\in\mathbf{C}]$$

subject to:

$$i_0 = n_0 + E\left(b_1\right)i_0 - \tau, \tag{25}$$

$$k_s^T F'\left(k_s^T\right) = -\left(x_s-b_{1s}\right)i_0 - \left(i_0-k_s^T\right)\theta \quad \forall s\text{: partial or no liquidation } (s\notin\mathbf{C}), \tag{28}$$

$$0 \le b_{1s} \le \theta F'\left(k_s^T\right) \qquad \forall s, \tag{29}$$

and:

$$E\left[3e+\tau+F\left(k^T\right)-F'\left(k^T\right)k^T-w\right] \ge U^{CE}. \tag{30}$$

To show that the competitive allocation is not constrained efficient, it is sufficient to show that the social planner can increase welfare by decreasing borrowing and investment in period 0. Such a change has several effects:

1. It reduces welfare by lowering the level of *ex ante* investment, i_0.

2. It increases welfare by reducing liabilities, b_{1s}, in certain states.

3. It reduces the amount of capital that has to be sold in fire sale states, increasing the asset price in those states.

4. It reduces the likelihood of a crisis.

We wish to determine when the net effect on welfare is positive. The positive contributions to welfare arise directly from the lower level of asset sales in fire sale states, and indirectly from a decrease in the likelihood of a crisis. We derive a condition under which the direct mechanism alone gives a positive net effect. Considering the indirect effect would strengthen our results but the analysis depends on the specific distributional assumptions taken and there is generally no closed-form solution.

Starting from the competitive allocation, suppose the social planner reduces *ex ante* investment by Δi_0 and reduces borrowing by the same amount against states in which $z_0 = z_s$ (z_0 and z_s are *ex ante* and *ex post* returns, as defined in Appendix A). First note that reducing borrowing against these states has no negative welfare effect on intermediaries since they are indifferent between investing *ex post* in them and *ex ante* in general. Therefore, to determine whether the reduction in i_0 is welfare-improving, we simply need to consider whether the welfare cost to consumers can be fully compensated for by any gain to intermediaries.

Differentiating the market clearing condition for used capital (which is obtained by equating supply, (15), and demand (17)), we can see that the reduction in i_0 decreases the amount of capital sold in 'recession' states by:

$$\frac{dk_s^T}{di_0} = \frac{x_s + \theta - b_{1s}}{[F'(k^T) - \theta] + F''(k^T)\,k^T}. \tag{31}$$

The profit consumers obtain from operating their technology is $F\left(k^T\right) - F'\left(k^T\right)k^T$. Therefore, in 'recession' states, the reduction in i_0 has a direct welfare cost to consumers of:

$$\rho_s = \frac{d\left[F\left(k^T\right) - F'\left(k^T\right)k^T\right]}{dk_s^T}\frac{dk_s^T}{di_0} = -\frac{x_s + \theta - b_{1s}}{[F'(k^T) - \theta] + F''(k^T)\,k^T}F''\left(k^T\right)k^T. \tag{32}$$

Intuitively, ρ_s represents the amount of goods transferred in 'recession' states from consumers to intermediaries as a result of the social planner's implementation of an equilibrium with lower borrowing than the competitive equilibrium. Intermediaries have to transfer at least this amount to consumers (in period 0, when they have resources to do so) to compensate them for this loss. What needs to be shown is that the net effect of this transfer is positive for intermediaries.

This will be the case if:

$$E(\rho) z_0 < E(\rho z). \tag{33}$$

The left hand side of (33) is the cost of the transfer to intermediaries and the right hand side is the benefit. In period 0, intermediaries transfer $E(\rho)$ goods to consumers, which they could have invested at a return z_0. On the other hand, intermediaries now have extra resources of ρ_s in each 'recession' state in period 1. Since returns on additional capital in period 1 are z_s, the expected benefit from these extra resources is $E(\rho z)$ .Without specifying the distribution of x and the parameter values, we cannot be specific about when this inequality is satisfied. However, provided that the distribution of x has sufficient variance, so that states in which $z_s > z_0$ are not very isolated events, it is generally satisfied (note that the positive correlation between ρ and z helps in this regard). If this is the case, welfare is unambiguously higher under the social planner's allocation than under the competitive equilibrium.

Appendix C: Implications of Changes in the Maximum Loan-to-Value Ratio

In this appendix, we show that increases in the maximum loan-to-value ratio, θ, heighten the scale of crises but have an ambiguous effect on their probability. Recall that we measure the likelihood of a crisis by $H(x^C) = \Pr[x < x^C]$ and its scale in terms of the asset price, q^C, which prevails in it.

We start by analysing the scale of crises. Substituting (16) into (17) gives the asset price in crises:

$$q^C = 1 - 2\alpha i_0. \tag{34}$$

In general, if θ increases, intermediaries can borrow more against those states in which they are constrained, which serves to increase their initial investment, i_0. There are only two channels through which intermediaries' investment could be reduced by an increase in θ. First, there is a general equilibrium channel by which an increase in θ may decrease the price of second hand capital in certain states, thus reducing the value of collateral in those states and, hence, reducing borrowing against those states. But this can only happen if, overall, initial investment has increased as a result of the increase in θ – as such, it can only ever be an offsetting channel. Second, an increase in θ may lower the likelihood of crises, which could reduce *ex ante* borrowing given that borrowing is positive against crisis states but may be zero against certain 'recession' states (see Figure 3). However, this effect has very little significance since crisis states are much rarer than states in which intermediaries are constrained. Given this, it follows that initial investment, i_0, is a positive function of θ. From (34), this implies that crises become more severe as the maximum loan-to-value ratio rises.

In terms of the probability of crisis, first note that from (24), the crisis threshold below which crises are unavoidable is given by:

$$x^C = -\left[\frac{(1-\theta)^2}{8\alpha i_0} + \theta\right]. \tag{24}$$

Differentiating with respect to θ gives:

$$\frac{\partial x^C}{\partial \theta} = \frac{\partial x^C}{\partial \theta} + \frac{\partial x^C}{\partial i_0}\frac{\partial i_0}{\partial \theta} \tag{35}$$

$$= \left[\frac{1-\theta}{4\alpha i_0} - 1\right] + \frac{(1-\theta)^2}{8\alpha i_0^2}\frac{\partial i_0}{\partial \theta}. \tag{36}$$

When $\theta = 1$, this expression is negative, implying that the crisis threshold is falling and crises becoming less likely as θ increases. So, in the vicinity of $\theta = 1$, it must be the case that increases in the maximum loan-to-value ratio reduce the probability of crises. The case where $\theta = 0$ is less clear cut as the sign of the first term in (24) is ambiguous. But, when $\theta = 0$, initial investment by intermediaries, i_0, is restricted to their initial net worth, n_0. Therefore, if initial net worth is sufficiently small, the first term in (36) is positive when $\theta = 0$, as is the whole expression, implying that the likelihood of crises is increasing in θ. So, increases in θ have an ambiguous effect on the probability of crises, serving to reduce their probability for high values of θ but generally increasing their probability for low values of θ.

References

Aghion, P., Bacchetta, P. and Banerjee, A. (2001). 'Currency crises and monetary policy in an economy with credit constraints', *European Economic Review*, vol. 45(7), pp. 1121-1150.

Aghion, P., Bacchetta, P. and Banerjee, A. (2004). 'Financial development and the instability of open economies', *Journal of Monetary Economics*, vol. 51(6), pp. 1077-1106.

Aghion, P., Banerjee, A. and Piketty, T. (1999). 'Dualism and macroeconomic volatility', *Quarterly Journal of Economics*, vol. 114(4), pp. 1359-1397.

Allen, F. and Carletti, E. (2006). 'Credit risk transfer and contagion', *Journal of Monetary Economics*, vol. 53(1), pp. 89-111.

Allen, F. and Gale, D. (2007). 'Systemic risk and regulation', In (M. Carey and R.M. Stulz, eds.), *The Risks of Financial Institutions*, pp. 341-374, Chicago: The University of Chicago Press.

Benati, L. (2004). 'Evolving post-World War II U.K. economic performance', *Journal of Money, Credit and Banking*, vol. 36(4), pp. 691-717.

Borio, C. (2004). 'Market distress and vanishing liquidity: anatomy and policy options', *Bank for International Settlements Working Paper*, Number 158.

Borio, C., Furfine, C. and Lowe, P. (2001). 'Procyclicality of the financial system and financial stability: issues and policy options', *Bank for International Settlements Papers*, Number 1.

Borio, C. and Lowe, P. (2002). 'Asset prices, financial and monetary stability: exploring the nexus', *Bank for International Settlements Working Paper*, Number 114.

Brunnermeier, M. K. and Pedersen, L. H. (2007). 'Market liquidity and funding liquidity', *NBER Working Paper*, Number 12939.

Caprio, G. and Klingebiel, D. (1996). 'Bank insolvencies: cross country experience', *World Bank Policy Research Working Paper*, Number 1620.

Coval, J. D. and Stafford, E. (2007). 'Asset fire sales (and purchases) in equity markets', *Journal of Financial Economics*, forthcoming.

Demirguc-Kunt, A. and Detragiache, E. (2005). 'Cross-country empirical studies of systemic bank distress: a survey', *IMF Working Paper*, Number 05/96.

Diamond, D. W. and Dybvig, P. H. (1983). 'Bank runs, deposit insurance, and liquidity', *Journal of Political Economy*, vol. 91(3), pp. 401-419.

Duffie, D., Garleanu, N. and Pedersen, L. H. (2005). 'Over-the-counter markets', *Econometrica*, vol. 73(6), pp. 1815-1847.

Eichengreen, B. and Mitchener, K. (2003). 'The Great Depression as a credit boom gone wrong', *Bank for International Settlements Working Paper*, Number 137.

Fisher, I. (1933). 'The debt-deflation theory of great depressions', *Econometrica*, vol. 1(4), pp. 337-357.

Gai, P., Jenkinson, N. and Kapadia, S. (2007). 'Systemic risk in modern financial systems: analytics and policy design', *Journal of Risk Finance*, vol. 8(2), pp. 156-165.

Gai, P., Kondor, P. and Vause, N. (2006). 'Procyclicality, collateral values, and financial stability', *Bank of England Working Paper*, Number 304.

Geithner, T. F. (2006). 'Risk management challenges in the U.S. financial system', Remarks at the Global Association of Risk Professionals 7th Annual Risk Management Convention and Exhibition, New York City, 28 February.

Gieve, J. (2006). 'Financial system risks in the United Kingdom – issues and challenges', *Bank of England Quarterly Bulletin*, vol. 46(3), pp. 337-341.

Hoggarth, G., Reis, R. and Saporta, V. (2002). 'Costs of banking system instability: some empirical evidence', *Journal of Banking and Finance*, vol. 26(5), pp. 825-855.

Holmstrom, B. and Tirole, J. (1998). 'Private and public supply of liquidity', *Journal of Political Economy*, vol. 106(1), pp. 1-40.

Jermann, U. and Quadrini, V. (2006). 'Financial innovations and macroeconomic volatility', *NBER Working Paper*, Number 12308.

Kehoe, T. J. and Levine, D. K. (1993). 'Debt-constrained asset markets', *Review of Economic Studies*, vol. 60 (October), pp. 865-888.

Kiyotaki, N. and Moore, J. (1997). 'Credit cycles', *Journal of Political Economy*, vol. 105(2), pp. 211-248.

Krishnamurthy, A. (2003). 'Collateral constraints and the amplification mechanism', *Journal of Economic Theory*, vol. 111(2), pp. 277-292.

Lorenzoni, G. (2005). 'Inefficient credit booms', Mimeo, MIT.

McGuire, P., Remolona, E., and Tsatsaronis, K. (2005). 'Time-varying exposures and leverage in hedge funds', *Bank for International Settlements Quarterly Review*, March, pp. 59-72.

Morris, S. and Shin, H. S. (1998). 'Unique equilibrium in a model of self-fulfilling currency attacks', *American Economic Review*, vol. 88(3), pp. 587-597.

Obstfeld, M. (1996). 'Models of currency crises with self-fulfilling features', *European Economic Review*, vol. 40(3-5), pp. 1037-1047.

Plantin, G., Sapra, H. and Shin, H. S. (2007). 'Marking-to-market: panacea or Pandora's box?', *Journal of Accounting Research*, forthcoming.

Pulvino, T. C. (1998). 'Do asset fire sales exist? An empirical investigation of commercial aircraft transactions', *Journal of Finance*, vol. 53(3), pp. 939-978.

Rajan, R. G. (2005). 'Has financial development made the world riskier?', Paper presented at the Federal Reserve Bank of Kansas City Economic Symposium on 'The Greenspan Era: Lessons for the Future', Jackson Hole, 25-27 August.

Shleifer, A. and Vishny, R. W. (1992). 'Liquidation values and debt capacity: a market equilibrium approach', *Journal of Finance*, vol. 47(4), pp. 1343-1366.

Stock, J. H. and Watson, M. W. (2005). 'Understanding changes in international business cycle dynamics', *Journal of the European Economic Association*, vol. 3(5), pp. 968-1006.

Tucker, P. (2005). 'Where are the risks?', *Bank of England Financial Stability Review*, vol. 19 (December), pp. 73-77.

White, W. R. (2004). 'Are changes in financial structure extending safety nets?', *Bank for International Settlements Working Paper*, Number 145.

Williamson, O. E. (1988). 'Corporate finance and corporate governance', *Journal of Finance*, vol. 43(3), pp. 567-591.

Monetary Policy and its Informative Value*

Romain Baeriswyl

Ludwig-Maximilians Universität München

Camille Cornand†

CNRS - BETA

First version: December 2005 - This version: October 2007

Abstract

This paper analyzes the welfare effects of economic transparency in a model of monopolistic competition with heterogeneous information on the shocks affecting the economy where the central bank has no inflationary bias. Monetary policy plays a dual role: a role as an action aiming at stabilizing the economy and a role as a public signal that partially reveals to firms the central bank's

*We thank Marios Angeletos and Christian Hellwig for their extensive feedbacks and guidance. For useful advice and comments at various stages of the project, we thank Laurent Clerc, Alex Cukierman, Etienne Farvaque, Petra Geraats, Charles Goodhart, Marvin Goodfriend, Frank Heinemann, Gerhard Illing, Hubert Kempf, Thomas Laubach, Andrew Levin, Olivier Loisel, Marc-Alexandre Sénégas, Hyun Shin, Eric Swanson and Carl Walsh. Remaining errors are the authors' own. The first version of this paper has been written when the second author was research officer at the London School of Economics; she thanks the FMG for hospitality and gratefully acknowledges financial support from the U.K. ESRC under grant RES-156-25-0026 ("Stability of the Global Financial System: Regulation and Policy Response", World Economy and Finance, ESRC Research Programme Phase I).

†Author in charge of the correspondence: BETA UMR 7522 CNRS - 61, avenue de la Forêt Noire - 67085 Strasbourg Cedex France - tel: +333 90 24 20 97 - fax: +333 90 24 20 71 - e-mail: cornand@cournot.u-strasbg.fr

assessment of the economy. Under opacity, firms are unable to decipher the central bank's assessment inducing its instrument: the central bank optimally balances the action and information purposes of its instrument. We derive the optimal monetary policy and central bank's disclosure strategy.

JEL classification: E52, E58, D82.

Keywords: heterogeneous information, monetary policy, signaling, transparency.

1 Introduction

This paper deals with the optimal conduct of monetary policy when imperfect information is responsible for money nonneutrality. Some scholars recently emphasized that, in an economy characterized by strategic complementarities, heterogeneous information can lead to realistic dynamics of transmission mechanisms.[1] Yet, the implications of the imperfect information hypothesis for the conduct of monetary policy remain largely unexplored. In our analysis, we emphasize the importance of central bank's communication in such an environment. Indeed, the communication strategy adopted by the central bank drives the real effect of monetary policy since it affects the degree of information imperfection and heterogeneity in the economy. Communication turns out to be an essential component in designing the optimal monetary policy pattern.

Central bank's communication has been largely discussed over the last decade. The main contribution of our approach is to distinguish two ways by which a central bank communicates with the private sector. On the one hand, it communicates explicitly by disclosing information *via* an announcement; this is the form of communication with which the literature deals generally. On the other hand, we highlight

[1]Adam (2007), Hellwig (2002), Mankiw and Reis (2002), and Woodford (2003a).

that a central bank also communicates – implicitly – with the private sector when implementing its monetary policy. While monetary policy analysis usually accounts for the action role of the instrument, we also consider its signaling role. Under imperfect information, the instrument of the central bank is not only an action that stabilizes the economy but also a signal that partially reveals to firms its own imperfect assessment of the fundamentals.[2] The signaling role of monetary policy has been well documented by Romer and Romer (2000). Using US data, they show that "*the Federal Reserve's actions signal its information*" and that "*commercial forecasters raise their expectations of inflation in response to contractionary Federal Reserve actions [...]*" (Romer and Romer (2000, p. 430)). Monetary policy entails a dual role, as an action and as a vehicle for information.

The fact to consider the informative value of the instrument set by the central bank allows to point out two intertwined implications for the conduct of monetary policy. The first implication is that the central bank chooses its instrument by optimally balancing its action and information purposes. To choose its instrument, the central bank should consider the information its policy contains. The second implication is related to transparency issues. Since the instrument of the central bank both influences the economy and reveals information to the public, the disclosure strategy must be chosen according to the instrument that is implemented. In particular, if the central bank wishes to withhold information from the markets, it should adjust the conduct of monetary policy in order to limit the revelation of such information through its policy choices. So a particularity of our model is that transparency and standard monetary policy issues are intertwined.

The literature – dealing with central banks transparency – mainly focuses on the impact of economic and political transparency of central banks in the Barro and

[2] Walsh (2006), and Walsh (2007) also share this characteristic.

Gordon (1983) framework.[3] Our paper ignores the potential contribution of transparency to deal with time-inconsistency. Instead, we focus on the effect of economic transparency on the optimal monetary policy when transmission mechanisms are driven by information imperfection. We concentrate on well-established and credible central banks and analyse the effect of economic transparency in the case where the central bank has no inflationary bias and where the private sector perfectly knows its preferences.[4] So the question of transparency deals with the provision of central bank's information to the private sector about its economic assessment. Recently, part of the literature dealing with coordination games has raised questions about the value of having central banks provide more and better information to the public. In their seminal beauty contest paper, Morris and Shin (2002) – emphasizing the relevance of strategic complementarities underlying most of macroeconomic aggregates – argue that, in an environment characterized by strategic complementarities and a lack of common knowledge, more accurate public information may be detrimental to welfare because public information is attributed too large a weight relative to its face value. Their argument has received a great deal of attention in the academic literature, the financial press[5], and central banks[6]. In a closely related work, Amato et al. (2002) interpret the model by Morris and Shin (2002) as a Lucas-Phelps islands economy in which firms try to second-guess the pricing strategies of their competitors. Challenging this result, Hellwig (2005) shows in a fully micro-founded model that more accurate public information about monetary shocks is always welfare increasing because it reduces price dispersion. Angeletos and Pavan (2006a) and Angeletos and Pavan (2006b) underline that these results are sensitive to the

[3]See Geraats (2002) for an overview.

[4]This is natural in the current context of central bank independence and historically – and durable – low levels of inflation, since the credibility problem of monetary policy seems to be a thing of the past (Cukierman (2002)).

[5]See The Economist (2004).

[6]See for example Kohn (2005) and Issing (2005).

extent to which coordination is socially valuable. While the literature in the vein of Morris and Shin addresses transparency issues when the only task of the central bank is to disclose or withhold information, the present paper provides a framework that simultaneously accounts for the action and communication of the central bank.

Our analysis is based on a model of monopolistic competition with heterogeneous information where two shocks affect the economy, namely demand and mark-up shocks. Both the central bank and firms are uncertain about the true state of the economy. In our set-up, an *opaque* central bank does not explicitly share its information about the state of the economy with firms *via* an announcement. When the economy is simultaneously hit by several types of shocks, firms are unable to properly interpret the monetary instrument as they cannot understand the rationale behind it.[7] For instance, the central bank may implement an expansionary instrument either because of a negative demand shock or because of a negative mark-up shock. This confusion reduces the informative value of the instrument on both fundamental shocks and on the beliefs of others about these shocks. By contrast, a *transparent* central bank discloses explicit information so that it reveals to firms its assessment of the fundamental shocks.

This paper analyzes the welfare effect of economic transparency, that is the extent to which the central bank should reveal to firms its own assessment of the fundamental shocks (namely demand and mark-up shocks). We derive the optimal monetary policy and optimal central bank's disclosure strategy. The welfare analysis of transparency is driven by three intertwined effects.

[7]This mechanism reminds us of that in the standard literature on economic and political transparency based on the inflation bias argument. In this literature, as in our paper, the private sector is unable to disentangle two possible rationales for the observable monetary instrument. However, when central bank's preferences are unknown, firms cannot disentangle whether the rationale for the monetary instrument is the change in central bank's preferences or the response to mark-up shocks. As a result, the central bank is inclined to respond more contractively to mark-up shocks to avoid firms believing its inflation bias has risen.

First, transparency has a positive *incentive effect* on the optimal monetary policy. As firms are unable to properly understand the reasons behind the instrument under opacity, the central bank balances the action and information purposes of its monetary instrument. This distorts its policy away from what would be optimal with respect to the action purpose only. By contrast, under transparency, the central bank chooses an instrument optimal from the perspective of its sole action purpose.

Second, transparency has a positive *uncertainty effect* with respect to demand shocks. Reducing the fundamental and strategic uncertainties about demand shocks is welfare increasing. This is because demand shocks can be neutralized by the policy implemented by the central bank. Even if central bank's information about demand shocks is noisy, transparency is welfare increasing since it reveals the influence of monetary policy on the economy and this is part of the fundamentals to which firms have to respond.[8]

Third, transparency has a negative *uncertainty effect* with respect to mark-up shocks. Mark-up shocks cannot be neutralized by the central bank as they create a trade-off between price level and output gap stabilization. Reducing the fundamental and strategic uncertainty about mark-up shocks owing to transparency is consequently detrimental to welfare since it exacerbates the response of each firm to mark-up shocks and increases the resulting loss.[9]

Overall, we show that transparency is welfare increasing (i) when the degree of strategic complementarities is low, (ii) when the economy is not too affected by mark-up shocks (relative to other shocks), (iii) when the central bank is more

[8]This result is similar to Hellwig (2005) but the rationale is different in the present paper. While in the former transparency is beneficial because it enhances socially beneficial coordination, in the latter the reason is linked to the action of the central bank.

[9]In a more general framework without monetary policy, Angeletos and Pavan (2006a) anticipated the fact that removing uncertainty about inefficiency shocks as mark-up shocks is welfare detrimental.

inclined towards price level rather than output gap stabilization, (iv) when firms have relatively precise private information, and (v) when the central bank has information that is relatively precise on demand shocks and relatively imprecise on mark-up shocks. Hence, our framework gives a rationale for the development of the economy over the last decades. Increasing transparency[10] seems appropriate in the current context of declining occurrence and amplitude of mark-up shocks and increasing inclination of central banks towards price stabilization.

The remaining of the paper is structured as follows. Section 2 outlines a monopolistic competition economy, in which firms' pricing decisions represent strategic complements. Section 3 considers a benchmark case under common information that recalls standard findings in monetary policy. Section 4 turns to the case of heterogeneous information and examines the optimal monetary policy and transparency. This section considers how announcements affect the optimal policy responses to demand and mark-up shocks and whether transparency is welfare increasing. Finally section 5 concludes.

2 The economy

The economy is populated by a representative household, a *continuum* of monopolistic competitive firms, and a central bank. We abstract here from the microfounded market interactions since they are very standard and focus on the optimal behavior of firms.[11] The timing of the game is the following. In the first stage, the economy is hit by two types of shocks, demand and mark-up shocks that are normally and

[10]The increase in transparency in the conduct of monetary policy in recent years is studied by Eijffinger and Geraats (2006) and Dincer and Eichengreen (2006).

[11]See for instance Adam (2007) for the derivation of the microfoundations.

independently distributed:

$$
\begin{aligned}
g &\sim N(0, \sigma_g^2) \\
u &\sim N(0, \sigma_u^2).
\end{aligned}
$$

Both the central bank and firms receive a private signal on them. In the second stage, the central bank chooses its policy, the monetary instrument, that determines the nominal aggregate demand up to the demand shocks g. The central bank may also disclose information about the rationale behind its policy. In the third stage, based on their information, firms simultaneously set their price.

2.1 Firms

The central equation of our model is given by the optimal pricing rule of firms. It is derived from an economy where the representative household consumes a composite good *à la* Dixit-Stiglitz where goods are imperfect substitutes. In such a context, the optimal price set by firm i is

$$p_i = \mathbb{E}_i[p + \xi c + u], \tag{1}$$

where $\mathbb{E}_i$ is the expectation operator of firm i conditional on its information, p the overall price level, c the output gap, and u the mark-up shock. The pricing rule (1) says that each firm sets its price according to both its own expectations about the real output gap and the mark-up shock, and its expectations about the overall price level. Per definition, the nominal aggregate demand deviation is the sum of deviations of the output gap and the price level: *i.e.* $y = c + p$. So, one can write the pricing rule as

$$p_i = \mathbb{E}_i[(1-\xi)p + \xi y + u]. \tag{2}$$

The parameter ξ captures the impact of the real output gap on prices (through wages). A large ξ means that the representative household is highly risk averse and that output gaps imply large variations in wages and thereby in prices. We qualify such an economy as weakly extensive. ξ also describes whether prices are strategic complements or substitutes. We assume that $0 < \xi < 1$, which implies that prices are strategic complements, meaning that firms tend to raise their price whenever they expect the others to do so. This assumption seems very natural and captures the concept of beauty contest introduced by Keynes: firms base their decision not only on their own expectations of fundamentals but also on the so-called higher-order expectations, *i.e.* expectations of the average expectations of fundamentals, up to an infinite number of iterations.

We will present the information available to firms in the subsequent section since it depends on the disclosure strategy adopted by the central bank.

2.2 The central bank

Based on its information, the central bank minimizes both the variability of the output gap c and that of the price level p owing to its monetary instrument I:

$$\min_I \mathbb{E}_{cb}[\lambda c^2 + p^2], \tag{3}$$

where λ is the weight assigned to output gap variability. The preferences of the central bank are common knowledge among firms. We will solve the minimization problem of the central bank as a Stackelberg game in which the central bank is the leader: to minimize its loss, the central bank takes into account firms' behavior by integrating their reaction function in its loss. This represents the case of full commitment. The monetary instrument implemented by the central bank is a linear combination of its signals on the shocks: $I = \nu_1 g_{cb} + \nu_2 u_{cb}$. ν_1 and ν_2 are the policy coefficients, and g_{cb} and u_{cb} stand for the central bank's signals on demand and mark-up shocks, respectively.

The information structure of the central bank is as follows. The central bank receives a signal on both the demand and the mark-up shocks in private. Each signal – or estimate – deviates from the true fundamental value by an error term that is normally distributed:

$$g_{cb} = g + \eta, \quad \text{with} \quad \eta \sim N(0, \sigma_\eta^2)$$
$$u_{cb} = u + \mu, \quad \text{with} \quad \mu \sim N(0, \sigma_\mu^2),$$

where η and μ are independently distributed.

We assume that the monetary instrument I implemented by the central bank partially determines the nominal aggregate demand. More precisely, the nominal aggregate demand y is the sum of the central bank's instrument I and the demand shock g, *i.e.* $y = I + g$. So, the pricing rule becomes

$$p_i = \mathbb{E}_i[(1 - \xi)p + \xi g + u + \xi I]. \tag{4}$$

3 Common information

Standard monetary policy analysis assumes that information is common knowledge among firms. While this paper deals with monetary policy under heterogeneous information, the current section derives, as a benchmark, the optimal monetary policy under common information.

When information is perfect and common to all firms, every firm sets the same price. The pricing rule (4) then simplifies to

$$p_i = p = I + g + \frac{1}{\xi}u.$$

Note that the impact of mark-up shocks u on the price level increases with the degree of strategic complementarities $1-\xi$. This arises because the weight assigned to mark-up shocks increases with the extensivity of the economy. As discussed above, when the economy is highly extensive (ξ small), firms assign a smaller weight to the nominal aggregate demand and a relatively larger one to mark-up shocks.

When the central bank has perfect information as well, its instrument simplifies to

$$I = \nu_1 g + \nu_2 u.$$

The resulting loss under perfect information is

$$L = \lambda\Big(-\frac{1}{\xi}u\Big)^2 + \Big[(1+\nu_1)g + (\frac{1}{\xi}+\nu_2)u\Big]^2,$$

and minimizing the unconditional expected loss yields the following optimal monetary policy:

$$\begin{aligned}\nu_1 &= -1\\ \nu_2 &= -\frac{1}{\xi}.\end{aligned}$$

The corresponding unconditional expected loss is a function of the variance of mark-up shocks and yields

$$\mathbb{E}(L) = \frac{\lambda}{\xi^2}\sigma_u^2.$$

The coefficient ν_1 indicates that the central bank perfectly offsets demand shocks. Since the monetary instrument is part of the nominal aggregate demand, the central bank is able to offset demand shocks. By closing the output gap, the central bank also gets rid of price deviations. So demand shocks are perfectly neutralized.

By contrast, mark-up shocks cannot be neutralized by the central bank as they create a trade-off between price level and output gap stabilization. Indeed, in the absence of any monetary policy action, a positive mark-up shock raises the price level and generates a negative output gap, both of amplitude $\frac{1}{\xi}u$. While price level stabilization calls for a contractionary policy, output gap stabilization requires an expansionary one. Under perfect and common information, the optimal monetary policy coefficient ν_2 states that the central bank lowers its instrument by $-\frac{1}{\xi}$ when the mark-up shock increases by one unit (*i.e.* contractionary policy). As the price level increases because of a positive mark-up shock, the central bank contracts the nominal aggregate demand so that the price level is completely stabilized (*i.e.* $p = 0$).[12] The resulting output gap is $c = -\frac{1}{\xi}u$.

4 Heterogeneous information

We now turn to the more realistic case where firms have heterogeneous information about the state of the economy. In this section, we derive the optimal monetary policy as a function of central bank transparency and then analyze the welfare effect

[12] The complete stabilization of the price level arises because of the absence of frictions under common knowledge. Since the monetary instrument is common knowledge among firms under common information, it has no real effect on output gap but only a nominal one on the price level.

of transparency. As information provided by the monetary instrument influences firms' reaction, the optimal policy varies according to the communication strategy adopted by the central bank.

We assume that the monetary instrument is perfectly observed by firms. This corresponds to the current practice of most central banks. By setting its instrument publicly, the central bank implicitly discloses a public signal to firms. However, without additional information, firms are unable to understand the central bank's assessment of the economy: since the central bank responds to two shocks, the monetary instrument does not allow firms to disentangle the rationale behind the implemented policy. This is the reason why many central banks, additionally to revealing the level of their instrument (*e.g.* the level of the overnight interest rate), explain their decision. A clear trend in this respect is the switch towards communication of the minutes of Monetary Policy Committee discussions. This section precisely aims at evaluating such communication strategies by considering whether the central bank should disclose additional information in the form of an explicit announcement that precisely reveals to the private sector its view about the state of the economy.

The central bank chooses its instrument to minimize (3). Since both fundamental shocks and both error terms are independently normally distributed, the optimal instrument rule of the central bank is a linear combination of its signals and can be written as

$$I = \nu_1(g + \eta) + \nu_2(u + \mu). \tag{5}$$

We first present the case where the central bank does not announce the rationale behind its instrument (opacity) and second the case where it reveals its own signals (transparency). Then we compare and discuss the optimal disclosure policy.

4.1 No announcement (opacity)

Each firm i receives a private signal on the mark-up shock u_i that may be interpreted as a private estimate. The private signal of each firm deviates from the true mark-up shock by an error term that is normally distributed:

$$u_i = u + \rho_i, \quad \text{with} \quad \rho_i \sim N(0, \sigma_\rho^2),$$

where ρ_i are identically and independently distributed across firms.

Firms also receive a public signal in the form of the monetary policy instrument (5). By setting its instrument, the central bank gives an indication to firms of its own beliefs about the state of the economy. Yet, without announcement, firms are uncertain about the right interpretation of the monetary instrument and about how others may interpret it. Firms rationally use the monetary instrument to infer the fundamental shocks g and u, and the expectations of other firms about these shocks.

4.1.1 Equilibrium

To determine the perfect Bayesian equilibrium behavior of firms, we recall the optimal pricing rule (4) for convenience and substitute successively the average price level with higher order expectations about demand and mark-up shocks and the monetary instrument:

$$\begin{aligned} p_i &= \mathbb{E}_i[(1-\xi)p + \xi g + u + \xi I] \\ &= \mathbb{E}_i\Big[\xi g + u + \xi I + (1-\xi)\Big[\bar{\mathbb{E}}[\xi g + u + \xi I + (1-\xi)[\bar{\mathbb{E}}[\xi g + u + \xi I + \ldots]]]\Big]\Big]. \end{aligned}$$

We denote by $\mathbb{E}_i(.)$ the expectation operator of firm i conditional on its information and by $\bar{\mathbb{E}}(.)$ the average expectation operator such that $\bar{\mathbb{E}}(.) = \int_i \mathbb{E}_i(.)di$. With heterogeneous information, the law of iterated expectations fails and expectations of higher order do not collapse to the average expectation of degree one.[13] Thus, we rewrite the pricing rule as

$$p_i = \sum_{k=0}^{\infty}(1-\xi)^k \mathbb{E}_i\Big[\bar{\mathbb{E}}^{(k)}(\xi g + u + \xi I)\Big],$$

and averaging over firms yields

$$p = \sum_{k=0}^{\infty}(1-\xi)^k \Big[\bar{\mathbb{E}}^{(k+1)}(\xi g + u + \xi I)\Big], \tag{6}$$

where k is the degree of higher order iterations. We use the notation: $\bar{\mathbb{E}}^{(0)}(x) = x, \bar{\mathbb{E}}^{(1)}(x) = \bar{\mathbb{E}}(x)$, and $\bar{\mathbb{E}}^{(2)}(x) = \bar{\mathbb{E}}\bar{\mathbb{E}}^{(1)}(x) = \bar{\mathbb{E}}\bar{\mathbb{E}}(x)$. The price level p is a weighted average of higher order expectations of the nominal aggregate demand $g + I$ and the mark-up shock u.[14] The corresponding output gap is given by

$$c = y - p = g + I - \sum_{k=0}^{\infty}(1-\xi)^k \Big[\bar{\mathbb{E}}^{(k+1)}(\xi g + u + \xi I)\Big].$$

The output gap is the difference between the nominal aggregate demand and the weighted average of higher order expectations of the demand shock g, the mark-up shock u, and the monetary instrument I. As fundamental and strategic uncertainties about nominal aggregate demand increase, the real effect of variations in demand increases as well. In the particular case where it is common knowledge, nominal aggregate demand has only a price effect.

In order to solve the inference problem of each firm

[13] See Morris and Shin (2002).

[14] An alternative and more intuitive way to formalize the price setting of firms would be to postulate that each firm has a known individual mark-up shock (instead of a signal on the aggregate mark-up shock u) as in Walsh (2007). Such an assumption would not qualitatively affect the results since they are mainly driven by strategic rather than fundamental uncertainty.

$$\mathbb{E}_i(g, u) = \mathbb{E}[g, u | u_i, I],$$

we define the corresponding covariance matrix $\mathbf{V}_{4\times 4}$ and the relevant sub-matrices

$$\mathbf{V} = \begin{pmatrix} \mathbf{V_{uu}} & \mathbf{V_{uo}} \\ \mathbf{V_{ou}} & \mathbf{V_{oo}} \end{pmatrix}.$$

The expectation of shocks conditional on the private and public signals of firm i is given by

$$\mathbb{E}\left(\begin{array}{c|c} g & \\ & u_i, I \\ u & \end{array} \right) = \mathbf{\Omega} \begin{pmatrix} u_i \\ I \end{pmatrix} = \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix} \begin{pmatrix} u_i \\ I \end{pmatrix},$$

where $\mathbf{\Omega} = \mathbf{V_{uo}} \mathbf{V_{oo}^{-1}}$.

Using this, equation (6) becomes

$$p = \sum_{k=0}^{\infty} (1-\xi)^k \Big[\begin{pmatrix} \xi & 1 \end{pmatrix} \mathbf{\Omega} \mathbf{\Xi^k} \begin{pmatrix} u \\ I \end{pmatrix} + \xi I \Big],$$

where

$$\mathbf{\Xi} = \begin{pmatrix} \Omega_{21} & \Omega_{22} \\ 0 & 1 \end{pmatrix}.$$

The equilibrium strategy for firm i is a linear combination of its private signal on mark-up shocks u_i and the public signal I:

$$p_i = \gamma_1 u_i + \gamma_2 I \qquad \text{with} \tag{7}$$

$$\gamma_1 = \frac{\xi \Omega_{11} + \Omega_{21}}{1 - (1-\xi)\Omega_{21}}$$

$$\gamma_2 = \frac{(1-\xi)\gamma_1\Omega_{22} + \xi(1+\Omega_{12}) + \Omega_{22}}{\xi}.$$

4.1.2 Optimal monetary policy

This section derives the optimal monetary policy under opacity. The central bank sets its monetary instrument (5) to minimize the expected loss (3) subject to the price rule (7). The unconditional expected loss is given by

$$\mathbb{E}(L) = \text{var}(p) + \lambda \cdot \text{var}(c).$$

First, the variance of the price level p can be written as

$$\text{var}(p) = (\gamma_2\nu_1)^2\sigma_g^2 + (\gamma_2\nu_1)^2\sigma_\eta^2 + (\gamma_1 + \gamma_2\nu_2)^2\sigma_u^2 + (\gamma_2\nu_2)^2\sigma_\mu^2.$$

Second, we determine the variance of the output gap. The output gap is

$$\begin{aligned} c &= I + g - p \\ &= g - \gamma_1 u + (1-\gamma_2)I. \end{aligned}$$

Therefore, the variance of the output gap yields

$$\begin{aligned} \text{var}(c) &= (1 + (1-\gamma_2)\nu_1)^2\sigma_g^2 + ((1-\gamma_2)\nu_1)^2\sigma_\eta^2 \\ &\quad + ((1-\gamma_2)\nu_2 - \gamma_1)^2\sigma_u^2 + ((1-\gamma_2)\nu_2)^2\sigma_\mu^2. \end{aligned}$$

As monetary policy is both an action and a vehicle for information, the central bank chooses its instrument by optimally balancing its action and information purposes.

The instrument that is optimal from the perspective of its action is given by the optimal monetary policy in the case where both the central bank and firms share the same information. Indeed, when firms already know (before observing the instrument) the assessment of the central bank about the state of the economy, the central bank has no incentive to distort its instrument in order to disguise its signals.

However, as soon as firms have imperfect information about the central bank's assessment of the state of the economy, the central bank can reduce its loss by considering also the informative value of its instrument. As we shall see below, transparency is welfare detrimental with respect to mark-up shocks while it is welfare improving with respect to demand shocks. As a result, the information purpose of monetary policy calls for making the instrument as less informative as possible on mark-up shocks.

Figure 1 shows the optimal monetary policy as a function of σ_ρ^2, the variance of the error terms of firms' private signal on mark-up shocks. The precision of firms' information declines moving from the left to the right part of the graph. The optimal monetary policy is computed with the following parameter values: $\sigma_g^2 = 1$, $\sigma_u^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\mu^2 = 0.2$, and $\lambda = 1$. Three cases can be distinguished with respect to the precision of firms' information.

First, when firms have perfect information on the mark-up shock ($\sigma_\rho^2 = 0$), the central bank implements the policy that is optimal from the perspective of its action and ignores the informative value of its instrument. Indeed, the central bank has no incentive to disguise its signal on the mark-up shock by altering its policy because firms already know the true mark-up shock. At the same time, revealing its signal on the demand shock to firms is not welfare detrimental since demand shocks can be neutralized. The strength of demand shock neutralization depends on the precision of central bank's information. In the present case where the variance of the error term is one fifth of the variance of the true demand shock, the optimal neutralization becomes $\nu_1 = -\frac{\sigma_g^2}{\sigma_g^2+\sigma_\eta^2} = -0.833$. In a similar way, the response of the central bank to mark-up shocks $\nu_2 = -\frac{1}{\xi}\frac{\sigma_u^2}{\sigma_u^2+\sigma_\mu^2}$ increases (in absolute value) with the precision of its information. The response to mark-up shocks also depends on the degree of strategic complementarities. As the latter increases, mark-up shocks are given an increasing relative weight in the pricing decision of firms and the central bank responds more strongly. With higher complementarities, monetary policy is less effective because nominal aggregate demand management has a small impact on prices.

Second, when firms' private information is extremely noisy, again the central bank fully neutralizes demand shocks according to the precision of its information, *i.e.* $\nu_1 \to -\frac{\sigma_g^2}{\sigma_g^2+\sigma_\eta^2}$ as $\sigma_\rho^2 \to \infty$. However, the central bank does not respond to mark-up shocks because firms do not respond to them since they get very noisy private signals, *i.e.* $\nu_2 \to 0$ as $\sigma_\rho^2 \to \infty$. Remember that the amplitude of the impact of the mark-up shock depends on the reaction of firms.

Third, for intermediate values of information precision, the optimal monetary policy depends on both the precision of private information and the degree of strategic complementarities. We first describe the central bank's response to mark-up shocks and then its response to demand shocks.

The optimal policy can be divided into two policy regions. When $\lambda\frac{\sigma_\rho^2}{\sigma_u^2} < \xi$, the central bank responds to mark-up shocks by contracting the nominal aggregate demand whenever its signal on the mark-up shock is positive (*i.e.* $\nu_2 < 0$). The strength of the policy response to mark-up shocks ν_2 declines with σ_ρ^2: as the quality of firms' information decreases, prices react also less to firms' expected mark-up shocks and the central bank finds it optimal to respond less strongly to them as well. But when $\xi < \lambda\frac{\sigma_\rho^2}{\sigma_u^2}$, it implements a slightly expansionary instrument (*i.e.*

$\nu_2 > 0$) whenever its signal on the mark-up shock is positive. The sign of the policy coefficient ν_2 depends on the effectiveness of monetary policy to stabilize the price level. Under opacity, the uncertainty of firms about the policy response of the central bank to mark-up shocks is large and this reduces the impact of the policy on the price level. As discussed in section 3, mark-up shocks create a trade-off between price level and output gap stabilization. The central bank is involved either in price level or output gap stabilization according to the effectiveness of its policy to stabilize the price level. This effectiveness is high when firms' fundamental and strategic uncertainty about the central bank's response to mark-up shocks is low. This arises either when firms' private information is highly accurate (*i.e.* private signals are good indicators for the central bank's response) or when strategic complementarities are weak (*i.e.* strategic uncertainty plays only a minor role). Otherwise, as uncertainty surrounding the response to mark-up shocks is high, the central bank finds it optimal to stabilize the output gap by expanding nominal demand in response to positive mark-up shocks. It is indeed more effective at stabilizing output gap rather than price level as achieving price level stabilization is more costly in terms of output gap.

The central bank always offsets demand shocks. But the amplitude of its response also depends on whether ξ is larger than $\lambda\frac{\sigma_\rho^2}{\sigma_u^2}$.

The central bank sets its response to demand shocks in order to reduce the informative value of its instrument about mark-up shocks. There are two ways for the central bank to achieve this goal. Either could the central bank weakly respond to demand shocks so that it avoids firms interpret the instrument as a response to mark-up shocks. Or could the central bank strongly respond to demand shocks so that firms mainly interpret the instrument as a response to demand shocks.

In the region where $\lambda\frac{\sigma_\rho^2}{\sigma_u^2} < \xi$, the central bank finds it optimal to respond more aggressively to demand shocks than it would do in the perspective of its sole action purpose. As firms have relatively precise information about mark-up shocks, the central bank strengthens its response to demand shocks to make its instrument less informative about mark-up shocks. Since the central bank strongly responds to mark-up shocks when firms' information is highly accurate, the central bank also strongly responds to demand shocks to mitigate the interpretation of its instrument. As explained above, this is precisely firms' reaction that determines the amplitude of the impact of the mark-up shock. By being less informative on the mark-up shock, the central bank limits the degree of common knowledge about the shock and therefore attenuates firms' reaction.

When $\lambda\frac{\sigma_\rho^2}{\sigma_u^2} = \xi$, as the central bank does not respond to mark-up shocks ($\nu_2 = 0$), the optimal response to demand shocks coincides with the policy required by a pure action motive.

And finally, when $\xi < \lambda\frac{\sigma_\rho^2}{\sigma_u^2}$, the central bank weakens its response to demand shocks. As firms' information about mark-up shocks is poorly accurate, the central bank finds it optimal to weakly respond to demand shocks so that firms do not interpret the instrument in a too large extent as a response to mark-up shocks. Compared to the policy case where the pure action purpose matters for the setting of the instrument, this policy reduces the informative value of the instrument about its mark-up shock signal.

4.2 Announcement (transparency)

Although the instrument provides information on the central bank's signals, it does not allow firms to properly understand the reason for the chosen monetary policy. As most central banks publish their instrument target, many of them are even more transparent and make the minutes of their Monetary Policy Committee deliberations available to the public. This reveals to the public the viewpoint of the central bank about the economy and rationalizes the monetary instrument.

As in the former case without announcement (opacity), each firm receives a private signal on the mark-up shocks u_i and the monetary instrument I is publicly available. With both demand and mark-up shocks hitting the economy, the sole observation of the monetary instrument does not allow firms to disentangle the extent to which each shock is responsible for the instrument. For example, the central bank may implement an expansionary instrument either because of a negative demand shock or because of a negative mark-up shock. In the current set-up, the central bank can remove uncertainty about the rationale for the instrument by explicitly announcing (one of) its signals. This renders the informative purpose of the monetary instrument ineffective and induces the central bank to implement its instrument for its action purpose only. We qualify such a central bank as *transparent* since its announcement eliminates any information asymmetry between itself and firms. For the sake of simplicity, we assume that the central bank directly announces its signal on the demand shock g_{cb}.[15] In this context, firms rationally use their three signals to infer the fundamental shocks and other firms' expectations about them.

4.2.1 Equilibrium

This section solves the perfect Bayesian equilibrium and derives the optimal behavior of firms and of the central bank. We proceed as in the former section to solve the inference problem each firm faces

$$\mathbb{E}[g, u, I | u_i, I, g_{cb}]$$

[15]One may think of different types of announcement that would reveal central bank's signals to firms. In practice, the publication of inflation forecast and/or target appears to be the main form of announcement adopted by transparent central banks.

and define the corresponding covariance matrix $\mathbf{V}_{\mathrm{T},6\times 6}$ and the relevant sub-matrices

$$\mathbf{V_T} = \begin{pmatrix} \mathbf{V_{T,uu}} & \mathbf{V_{T,uo}} \\ \mathbf{V_{T,ou}} & \mathbf{V_{T,oo}} \end{pmatrix}.$$

The expectation of the fundamental shocks conditional on the private and public signals of firm i is given by

$$\mathbb{E}\left(\begin{array}{c|c} g & \\ u & u_i, I, g_{cb} \\ I & \end{array}\right) = \mathbf{\Omega_T}\begin{pmatrix} u_i \\ I \\ g_{cb} \end{pmatrix} = \begin{pmatrix} \Omega_{\mathrm{T},11} & \Omega_{\mathrm{T},12} & \Omega_{\mathrm{T},13} \\ \Omega_{\mathrm{T},21} & \Omega_{\mathrm{T},22} & \Omega_{\mathrm{T},23} \\ 0 & 1 & 0 \end{pmatrix}\begin{pmatrix} u_i \\ I \\ g_{cb} \end{pmatrix},$$

where $\mathbf{\Omega_T} = \mathbf{V_{T,uo}}\mathbf{V^{-1}_{T,oo}}$.

Using this result into the price rule (6) yields

$$p = \sum_{k=0}^{\infty}(1-\xi)^k \begin{pmatrix} \xi & 1 & \xi \end{pmatrix} \mathbf{\Omega_T}\mathbf{\Xi^k_T}\begin{pmatrix} u \\ I \\ g_{cb} \end{pmatrix}, \tag{8}$$

where

$$\Xi_{\mathrm{T}} = \begin{pmatrix} \Omega_{\mathrm{T},21} & \Omega_{\mathrm{T},22} & \Omega_{\mathrm{T},23} \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

The price level equation (8) is a linear combination of the mark-up shock u and of the public signals I and g_{cb}:

$$p = \gamma_1 u + \gamma_2 I + \gamma_3 g_{cb} \qquad \text{with} \tag{9}$$

$$\begin{aligned}
\gamma_1 &= \frac{\xi\Omega_{\mathrm{T},11} + \Omega_{\mathrm{T},21}}{1-(1-\xi)\Omega_{\mathrm{T},21}} \\
\gamma_2 &= \frac{(1-\xi)\gamma_1\Omega_{\mathrm{T},22} + \xi(1+\Omega_{\mathrm{T},12}) + \Omega_{\mathrm{T},22}}{\xi} \\
\gamma_3 &= \frac{(1-\xi)\gamma_1\Omega_{\mathrm{T},23} + \xi\Omega_{\mathrm{T},13} + \Omega_{\mathrm{T},23}}{\xi}.
\end{aligned}$$

4.2.2 Optimal monetary policy

The central bank sets its monetary instrument to minimize the expected loss given the precision of its information. First, the variance of the price level p can be written as

$$\mathrm{var}(p) = (\gamma_2\nu_1 + \gamma_3)^2\sigma_g^2 + (\gamma_2\nu_1 + \gamma_3)^2\sigma_\eta^2 + (\gamma_1 + \gamma_2\nu_2)^2\sigma_u^2 + (\gamma_2\nu_2)^2\sigma_\mu^2.$$

Second, we determine the variance of the output gap. The output gap is

$$\begin{aligned}
c &= I + g - p \\
&= g - \gamma_1 u + (1-\gamma_2)I - \gamma_3 g_{cb}.
\end{aligned}$$

Therefore,

$$\begin{aligned}
\mathrm{var}(c) &= (1 + (1-\gamma_2)\nu_1 - \gamma_3)^2\sigma_g^2 + ((1-\gamma_2)\nu_1 - \gamma_3)^2\sigma_\eta^2 \\
&\quad + ((1-\gamma_2)\nu_2 - \gamma_1)^2\sigma_u^2 + ((1-\gamma_2)\nu_2)^2\sigma_\mu^2.
\end{aligned}$$

With the additional announcement, firms are able to perfectly disentangle the signals of the central bank. Thus the central bank cannot influence firms' beliefs by altering its instrument. The central bank does not face, unlike under opacity, the problem of optimally balancing the action and information purposes of its monetary instrument anymore. On the contrary, the central bank implements the instrument that is optimal from the perspective of its action purpose only. The corresponding coefficients of monetary policy satisfy:

$$\nu_1 = -\frac{\sigma_g^2}{\sigma_g^2 + \sigma_\eta^2} \quad (10)$$

$$\nu_2 = -\frac{1}{\xi}\frac{\sigma_u^2}{\sigma_u^2 + \sigma_\mu^2}. \quad (11)$$

As stated above, equation (10) indicates that the central bank tries to fully neutralize demand shocks according to the precision of its signal. The central bank's response to mark-up shocks (11) increases with the precision of its information. However, the response also depends on the degree of strategic complementarities since monetary policy is less effective for influencing the price level when the economy is highly extensive.

4.3 Welfare effect of transparency

This section analyzes the welfare effect of transparency. The main results are the following. First, transparency is welfare increasing with respect to demand shocks but detrimental with respect to mark-up shocks. As demand shocks can be neutralized by the central bank, reducing uncertainty about how the central bank responds to them helps stabilizing the economy. By contrast, reducing uncertainty about mark-up shocks is detrimental as it exacerbates firms' reaction and raises the resulting loss since the central bank cannot neutralize them. Transparency is welfare improving either when mark-up shocks are not too relevant compared to demand shocks or when the degree of strategic complementarities is low as firms' pricing decision relies less on mark-up shocks. Second, transparency is particularly beneficial when the central bank is more inclined towards price stabilization. Indeed, transparency increases the effectiveness of monetary policy on the price level.

We first describe the three mechanisms that drive these results. Then, we compare the welfare level under opacity *versus* transparency, and emphasize the impact of the degree of strategic complementarities $(1-\xi)$, of the precision of firms' private information σ_ρ^2, of the variance of mark-up shocks σ_u^2, and of the preference of the central bank for output gap stabilization λ.

4.3.1 Effects at stake

Our results are driven by three effects. First, transparency has a positive *incentive effect* on the optimal monetary policy. In the absence of transparency, firms are unable to disentangle the reasons behind the monetary instrument. Monetary policy then entails a dual role, which induces the central bank to optimally balance the action and information purposes of its instrument. Transparency eliminates the informative value of the instrument (or makes it redundant) and the central bank focuses on its action purpose. The incentive effect of transparency is welfare increasing as transparency allows the central bank to choose the instrument that optimally stabilizes the economy.

Second, transparency has a positive *uncertainty effect* with respect to demand shocks on the behavior of firms. Transparency reduces both fundamental and strategic uncertainties about demand shocks and central bank's response to them. Reducing this uncertainty is welfare improving since demand shocks can be neutralized by the central bank. Transparency reduces the distorting effect of a monetary instrument implemented by a central bank with poorly accurate information. This mainly departs from the conclusion by Morris and Shin (2002) because our framework additionally accounts for the action taken by the central bank.

Third, transparency has a negative *uncertainty effect* with respect to mark-up shocks. As mark-up shocks create a trade-off between price and output gap stabilization, they cannot be neutralized by the central bank. Reducing uncertainty about mark-up shocks is thus welfare detrimental because it exacerbates the reaction of firms to them. When strategic complementarities are strong, firms put a large weight on higher order expectations on mark-up shocks. Transparency reduces higher order uncertainty and induces firms to strongly react to mark-up shocks.

4.3.2 Degree of strategic complementarities and precision of private information

Figure 2 represents the ratio of the unconditional expected loss under transparency (*i.e.* with announcement) to the unconditional expected loss under opacity (*i.e.* without announcement) $\mathbb{E}(L_T)/\mathbb{E}(L_O)$ as a function of strategic complementarities ξ for three values of precision of firms' information σ_ρ^2. Transparency is welfare detrimental whenever the ratio is larger than one. The model is solved numerically with the following parameter values: $\sigma_g^2 = 1$, $\sigma_u^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\mu^2 = 0.2$, and $\lambda = 1$.

Transparency is welfare detrimental when the negative uncertainty effect with respect to mark-up shocks dominates both the positive incentive effect and the uncertainty effect with respect to demand shocks. Removing uncertainty about mark-up shocks is highly detrimental either when higher order expectations are given a large weight or when firms have very noisy information about them.

Figure 2 shows that transparency is welfare detrimental when the degree of strategic complementarities $(1 - \xi)$ is high. Price setting in an economy with a high degree of strategic complementarities is characterized by two intertwined features. First, prices are mainly determined by mark-up shocks when complementarities are high because demand shocks have a limited impact on prices as the economy is highly extensive. Second, firms are more sensitive to other firms' pricing decision. This implies that, with increasing strategic complementarities, firms put an increasing weight on higher order expectations on mark-up shocks. In this context, the detrimental effect of transparency is driven by the negative uncertainty effect related to mark-up shocks. Indeed, when strategic complementarities are strong, transparency, by reducing higher order uncertainty, induces firms to strongly react to mark-up shocks.

The precision of firms' private information strongly influences the effects at stake. In the case where firms' private information is very noisy, the detrimental uncertainty effect of transparency dominates its positive incentive effect. But when firms already have precise private information, reducing uncertainty on fundamental shocks and higher order expectations has a relatively small negative effect compared to the positive incentive effect. So, transparency is welfare detrimental when complementarities are high and as long as firms' private information is not too precise.

4.3.3 Relative importance of mark-up shocks

Figures 3 and 4 represent the ratio $\mathbb{E}(L_T)/\mathbb{E}(L_O)$ as a function of the variance of mark-up shocks for three levels of strategic complementarities. Other parameter values are $\sigma_g^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\mu^2 = 0.2\sigma_u^2$, $\sigma_\rho^2 = 0.2\sigma_u^2$, and $\lambda = 1$.

The variance of mark-up shocks σ_u^2 captures the importance of mark-up shocks in the economy. When there is no mark-up shock ($\sigma_u^2 = 0$), the question of transparency is irrelevant to welfare whatever the degree of strategic complementarities. As the central bank exclusively responds to demand shocks, firms perfectly interpret the rationale behind the monetary instrument even under opacity. So, the optimal monetary policy and the economic outcome cannot be distinguished between opacity and transparency.

However, as soon as σ_u^2 increases, figure 3 shows that the welfare effect of transparency depends on both the degree of strategic complementarities and the importance of mark-up shocks in the economy, relative to demand shocks. As discussed in the previous section, transparency tends to improve welfare when complementarities are weak. But whatever the degree of strategic complementarities, transparency turns out to be welfare detrimental as the relative importance of mark-up shocks increases. Indeed, since mark-up shocks cannot be neutralized by the central bank, the detrimental uncertainty effect of transparency dominates as mark-up shocks become more relevant. Figure 4 allows the variance of mark-up shocks to become very large. Transparency is welfare detrimental even in the case of low complementarities ($\xi = 0.7$) when the importance of mark-up shocks is very high relative to that of demand shocks.

4.3.4 Central bank's preference for output gap stabilization

Figure 5 illustrates the ratio $\mathbb{E}(L_T)/\mathbb{E}(L_O)$ as a function of σ_u^2 for three levels of λ, the weight the central bank assigns to output gap variability. The parameter values used for the simulation are $\sigma_g^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\mu^2 = 0.2\sigma_u^2$, $\sigma_\rho^2 = 0.2\sigma_u^2$, and $\xi = 0.5$.

It turns out that transparency is welfare improving when the central bank is more inclined towards price stabilization. Indeed, the central bank more effectively influences firms' behavior and thus the price level when it is transparent. As the central bank becomes more inclined towards price level stabilization (λ falls), the optimal policy coefficient in response to mark-up shocks ν_2 under opacity becomes more negative. As the central bank's influence on firms' behavior is limited by opacity, it finds it optimal to respond more strongly to shocks to better control the price level. This makes the monetary instrument more informative about mark-up shocks and considerably reduces the negative uncertainty effect of transparency.

Central banks are more inclined towards price stability today than they were in the past. Indeed, the recent switch from secrecy to transparency is clearly motivated by the will of central banks to publicly reveal their intention to stabilize prices.[16] In this respect, our model suggests that stronger price stabilization calls for higher economic transparency. Since the main aim of political transparency is better price stabilization, our result highlights that economic transparency should go along with political transparency.

4.3.5 Precision of central bank's signal on mark-up shocks

Figure 6 illustrates the ratio $\mathbb{E}(L_T)/\mathbb{E}(L_O)$ as a function of the precision of central bank's information on mark-up shocks σ_μ^2 for three levels of ξ. The parameter values used for the simulation are $\sigma_g^2 = 1$, $\sigma_u^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\rho^2 = 0.2$, and $\lambda = 1$.

This figure shows that transparency is welfare improving as the precision of central bank's signal on mark-up shocks decreases. The intuition is straightforward. Transparency is welfare detrimental when it exacerbates firms' reaction to mark-up shocks. But with poorly accurate central bank's information about mark-up shocks, the announcement does not contain much valuable information about them. As more accurate information on mark-up shocks exacerbates firms' reaction, noisy central bank's information reduces the pertinence of the announcement with respect to mark-up shocks. But, as transparency does not provide much information about mark-up shocks when σ_μ^2 is large, it provides firms with valuable information about demand shocks and central bank's response to them. The announcement however reveals to firms how the central bank perceives and responds to demand shocks, and reduces uncertainty about them.

[16]See Geraats (2002) and Rogoff (2003).

4.4 The relevance of the informative value of monetary policy

As we argue in this paper, it is optimal for a central bank to choose its policy by accounting for the information it conveys to the private sector. In particular, we have shown that the optimal monetary policy under opacity deviates from that under transparency since the central bank distorts its policy in order not to reveal some of its information. In this section, we address the relevance of considering the informative value for the central bank when it chooses its policy.

First, we compare the welfare under opacity when the central bank optimally accounts for the informative value of its policy (former section 4.3) to the case where it naively ignores that firms use its policy to infer fundamentals. In this case, the central bank sets its policy as if firms did not make use of it to improve their expectations. This exercise helps further appreciate the importance of adjusting the policy when the central bank wishes to withhold information. Not surprisingly, our analysis shows that it is always better to account for the informative value of monetary policy and that being unaware of it may yield large additional losses.

Second, we analyze the extent to which opacity may be welfare beneficial relative to transparency whenever the central bank ignores the informative value of monetary policy. It turns out that opacity is always welfare detrimental as soon as the informative value of policy is not accounted for. In other words, optimally balancing the action and information purposes of policy is a necessary condition for opacity to be welfare improving. This analysis emphasizes the relevance of the informative value of policy while addressing communication issues.

The naive policy under opacity, that ignores the signaling role of monetary policy, is such that the central bank believes firms know the monetary instrument I but do not use it in their bayesian update to revise their price. The price p_i set by a naive firm is therefore equal to

$$p_i = \frac{\sigma_u^2}{\sigma_\rho^2 + \xi\sigma_u^2} u_i + I.$$

Proceeding as earlier, the naive monetary policy is thus given by the following monetary policy coefficients:

$$\begin{aligned} \nu_1 &= 0 \\ \nu_2 &= -\frac{\sigma_u^4}{(\sigma_\rho^2 + \xi\sigma_u^2)(\sigma_\mu^2 + \sigma_u^2)}. \end{aligned}$$

To conduct our welfare analysis and to derive the naive policy under opacity, we plug those policy coefficients into the loss function of the central bank under the optimal opacity case. This means that firms make use of the information conveyed by the instrument, while the central bank ignores this fact.

4.4.1 Naive opacity *versus* optimal opacity

We compare the welfare under the truly optimal policy under opacity to the welfare under the naive policy under opacity for different parameter values and analyse particularly the impact of the variance of mark-up shocks. Trivially, we find that ignoring the fact that firms react to the signaling role of the instrument is costly in terms of welfare. The usefulness of the exercise lies in the fact that depending on parameter values, being naive for the central bank may be more or less detrimental.

Figure 7 illustrates this (trivial) result. It represents the ratio of the unconditional expected loss under the naive opacity case to the unconditional expected loss under the optimal opacity case $\mathbb{E}(L_N)/\mathbb{E}(L_O)$ as a function of the variance of mark-up shocks σ_u^2. The naive opacity is welfare detrimental whenever the ratio is larger than one. Parameter values are $\sigma_g^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\mu^2 = 0.2\sigma_u^2$, $\sigma_\rho^2 = 0.2\sigma_u^2$, $\xi = 0.5$, and $\lambda = 1$.

The ratio is always larger than one. But as the variance of mark-up shocks σ_u^2 increases (mark-up shocks become more relevant in the economy), being naive for the central bank becomes relatively less and less detrimental. The reason is that the effectiveness of hidding mark-up shocks owing to the informative device of the instrument is strong as long as mark-up shocks are not too relevant in the economy and as long as firms are not already too much informed about these shocks. By being naively opaque, the central bank partly reveals mark-up shocks to firms in a case where it could easily hide them by optimally distorting its instrument (by being optimally opaque), that is when mark-up shocks are not too relevant; this exacerbates the reaction of firms to mark-up shocks. When mark-up shocks become more relevant, the marginal information given by the naive instrument does not deteriorate welfare in a large extent.

Additional simulations show that when the precision of central bank's information on mark-up shocks is high, it is extremely costly in terms of welfare for the central bank to be naive. As the precision of central bank's information decreases, being naive tends to yield the same welfare as being optimal. The intuition is that when the central bank has precise information on mark-up shocks, not accounting for the informative value of monetary policy is costly because the instrument contains precise information. As the precision of central bank's information decreases, the informative value of its instrument in the case of an optimal monetary policy decreases as well.

4.4.2 Naive opacity *versus* optimal transparency

We now compare on figure 8 the relative losses $\mathbb{E}(L_T)/\mathbb{E}(L_O)$ and $\mathbb{E}(L_T)/\mathbb{E}(L_N)$ for two values of ξ and analyse the impact of the variance of mark-up shocks. Parameter values are $\sigma_g^2 = 1$, $\sigma_\eta^2 = 0.2$, $\sigma_\mu^2 = 0.2\sigma_u^2$, $\sigma_\rho^2 = 0.2\sigma_u^2$, and $\lambda = 1$.

The main result is that limiting transparency is welfare improving only if the central bank takes into account the informative value of its action. Indeed, welfare with the naive policy under opacity is always lower than welfare under optimal transparency, even in cases where optimal opacity is much better than optimal transparency.

When mark-up shocks are not too relevant, being naive is extremely detrimental. But as mark-up shocks become more relevant, being naively opaque tends to yield the same welfare as being optimally transparent. This result is again in line with our former standard analysis in figure 3. As mark-up shocks become more relevant in the economy, transparency generally tends to be more welfare detrimental. The reason is that the negative uncertainty effect of transparency with respect to mark-up shocks dominates both uncertainty and incentive positive effects as mark-up shocks become more relevant. And the informative value of the instrument in the naive opacity case is not very strong as mark-up shocks become more relevant (firms already know mark-up shocks and it is difficult for the central bank to hide them by distorting its instrument because the instrument contains almost only information about u). So the marginal information given by the naive instrument does not deteriorate welfare in a large extent.

4.5 Extension to welfare criterion incorporating the welfare cost of dispersion

So far, we have examined welfare along the lines of the standard objective function commonly used in the literature and which is made of two terms: c^2 which stands for the aggregate volatility of output and p^2 which stands in for inflation. While an objective function in this form has longly been intuitively rationalized, Woodford (2003b) has recently shown, within a fully micro-founded new Keynesian model of monopolistic competition, that the welfare of a representative household decreases both with the deviation of the output from its steady state and with the cross-sectional price dispersion.

In this section, we add price dispersion in our welfare criterion[17] and examine how this weight affects our results. As explained by Angeletos and Pavan (2006a), Hellwig (2005), Roca (2005), and Lorenzoni (2005), the application of the Morris and Shin argument to different welfare functions may lead to different conclusions. In particular, Hellwig (2005) and Woodford (2005) show that when coordination is socially highly valuable, transparency is welfare improving as it helps coordinating firms' price setting. In their model, the potential destabilizing effect of transparency is neglected.

The new loss function incorporating the welfare cost of dispersion when all terms are weighted equally is written as follows

$$L = c^2 + p^2 + \int_i (p_i - p)^2 di. \tag{12}$$

The economy remains the same as earlier (section 2.1).

[17]For the sake of simplicity, we only consider here the case were output gap deviation, price deviation and price dispersion are weighted equally.

With the new welfare function accounting for price dispersion (12), the unconditional expected loss is

$$\mathbb{E}(L) = \text{var}(c) + \text{var}(p) + (\gamma_1^2 \sigma_\rho^2), \tag{13}$$

where var(c) and var(p) are the same as in section 4.2.2 in the transparency case and var(c) and var(p) are the same as in section 4.1.2 in the opacity case.

Overall, simulations show that when there is a coordination motive in the welfare function, as expected, transparency becomes a bit more beneficial. When all weights are equal, simulations are very close to our standard case where there is no price dispersion in the loss function ($\lambda_3 = 0$). This result can easily be seen from figures 9, 10 and 11 (which are respectively close to figures 2, 3 and 6).

5 Concluding remarks

This paper analyzes the welfare effects of economic transparency in the conduct of monetary policy with heterogeneous information on the state of the economy. The main characteristic of our paper is to recognize that monetary policy entails a dual role: the instrument of the central bank is both an action that stabilizes the economy and a signal that partially reveals to firms the central bank's assessment of the state of the economy. We derive both the optimal monetary policy and the optimal central bank's disclosure.

The notion of transparency considered in this paper is the following. The observation of the monetary instrument does not allow firms to disentangle the central bank's opinion about each shock. A transparent central bank removes this uncertainty by disclosing an additional announcement that explains to the private sector the rationale behind its instrument. Under opacity, firms are unable to perfectly disentangle the central bank's signals responsible for the instrument. So, the central bank chooses its instrument by optimally balancing its action and information purposes. By contrast, under transparency, the central bank allows firms to identify the rationale behind the instrument and implements the policy that is optimal in the perspective of its sole action purpose.

In this context, we show that transparency is welfare increasing (i) when the degree of strategic complementarities is low, (ii) when the economy is not too affected by mark-up shocks, (iii) when the central bank is more inclined towards price stabilization, (iv) when firms have relatively precise private information, and (v) when the central bank has information that is relatively precise on demand shocks and relatively imprecise on mark-up shocks.

This result rationalizes the increase in central bank's transparency in the current context where mark-up shocks have a relatively low impact on the economic development. Since central banks that assign a large weight on price stabilization tend to be transparent with respect to their political targets, our framework suggests that economic transparency should go along with political transparency.

References

Adam, K. (2007). Optimal monetary policy with imperfect common knowledge. *Journal of Monetary Economics*, 54(2):276–301.

Amato, J., Morris, S., and Shin, H. S. (2002). Communication and monetary policy. *Oxford Review of Economic Policy*, 18(4):495–503.

Angeletos, G.-M. and Pavan, A. (2006a). Efficient use of information and social value of information. *Econometrica*, forthcoming.

Angeletos, G.-M. and Pavan, A. (2006b). Socially optimal coordination: Characterization and policy implications. *Journal of the European Economic Association*, forthcoming.

Barro, R. J. and Gordon, D. B. (1983). A positive theory of monetary policy in a natural rate model. *Journal of Political Economy*, 91(4):589–610.

Cukierman, A. (2002). Are contemporary central banks transparent about economic models and objectives and what difference does it make? *Federal Reserve Bank of St. Louis Review*, 84(4):14–45.

Dincer, N. and Eichengreen, B. (2006). Central bank transparency: Where, why, and to what effect? *mimeo*, http:// www.centrecournot.org/ pdf/ conference9/ Barry%20EICHENGREEN.pdf.

Eijffinger, S. C. W. and Geraats, P. M. (2006). How transparent are central banks? *European Journal of Political Economy*, 22(1):1–21.

Geraats, P. M. (2002). Central bank transparency. *The Economic Journal*, 112:F532–F565.

Hellwig, C. (2002). Public announcements, adjustment delays and the business cycle. *mimeo*, http://www.econ.ucla.edu/ people/papers/ Hellwig/Hellwig208.pdf.

Hellwig, C. (2005). Heterogeneous information and the welfare effects of public information disclosures. *mimeo*, http://www.econ.ucla.edu/ people/papers/ Hellwig/Hellwig283.pdf.

Issing, O. (2005). Communication, transparency, accountability: Monetary policy in the twenty-first century. *Federal Reserve Bank of St. Louis Review*, 87(2, pt.1):65–83.

Kohn, D. L. (2005). Central bank communication. *mimeo*, http://www.federalreserve. gov/boarddocs/Speeches/ 2005/20050109/ default.htm.

Lorenzoni, G. (2005). Imperfect information, consumers' expectations and business cycles. *mimeo*.

Mankiw, N. G. and Reis, R. (2002). Sticky information versus sticky prices: A proposal to replace the new keynesian phillips curve. *The Quarterly Journal of Economics*, 117:1295–1328.

Morris, S. and Shin, H. S. (2002). Social value of public information. *American Economic Review*, 92(5):1521–1534.

Roca, M. (2005). Transparency and monetary policy with imperfect common knowledge. *mimeo*.

Rogoff, K. (2003). Globalization and global disinflation. *mimeo*, http://www.kc.frb.org/ Publicat/sympos/ 2003/pdf/ Rogoff.0910.2003.pdf.

Romer, C. D. and Romer, D. H. (2000). Federal reserve information and the behavior of interest rates. *American Economic Review*, 90(3):429–457.

The Economist (2004). It's not always good to talk. July 22nd:71.

Walsh, C. E. (2006). Transparency, flexibility, and inflation targeting. In Mishkin, F. and Schmidt-Hebbel, K., editors, *Monetary Policy under Inflation Targeting.* Central Bank of Chile, Santiago, Chile.

Walsh, C. E. (2007). Optimal economic transparency. *International Journal of Central Banking*, 3(1):5–36.

Woodford, M. (2003a). Imperfect common knowledge and the effects of monetary policy. In Aghion, P., Frydman, R., Stiglitz, J., and Woodford, M., editors, *Knowledge, Information and Expectations in Modern Macroeconomics: In Honor of Edmund S. Phelps.* Princeton University Press, Princeton.

Woodford, M. (2003b). *Interest and Prices: Foundations of a Theory of Monetary Policy.* Princeton University Press.

Woodford, M. (2005). Central bank communication and policy effectiveness. *mimeo*, http://www.columbia.edu/ %7Emw2230/JHole05.pdf.

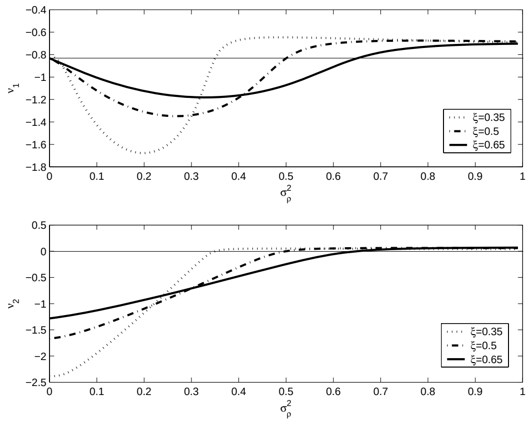


Figure 1: Optimal monetary policy under opacity

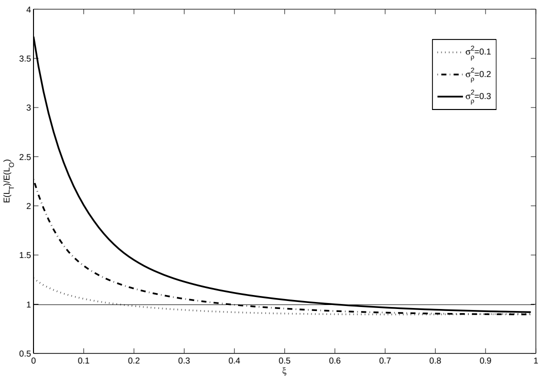


Figure 2: Welfare effect of transparency: impact of ξ

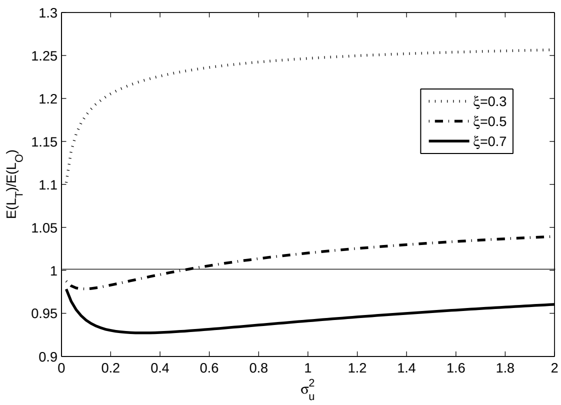


Figure 3: Welfare effect of transparency: impact of σ_u^2

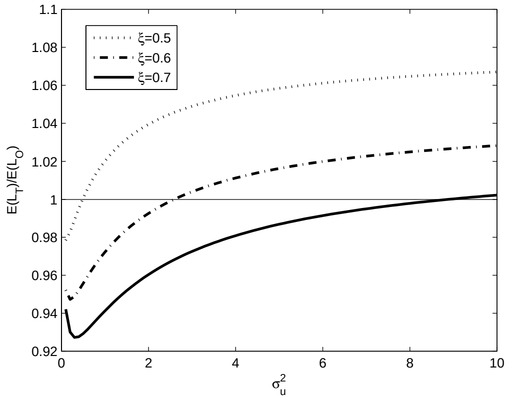


Figure 4: Welfare effect of transparency: impact of σ_u^2

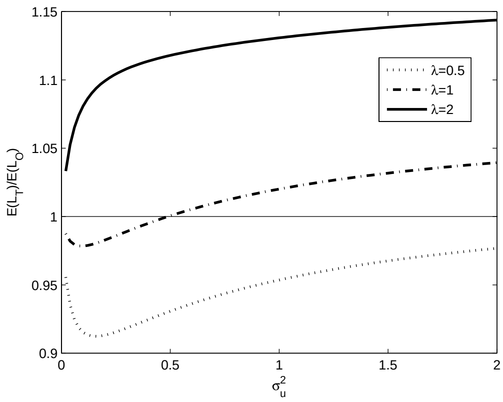


Figure 5: Welfare effect of transparency: impact of λ

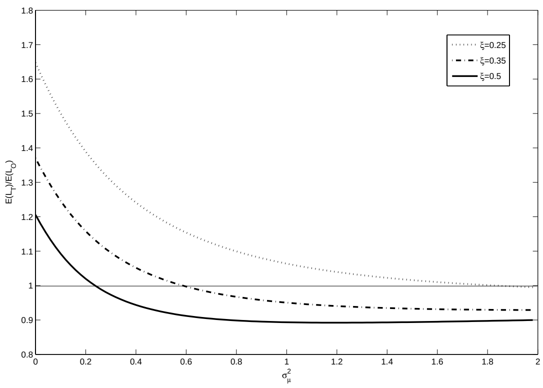


Figure 6: Welfare effect of transparency: impact of σ_μ^2

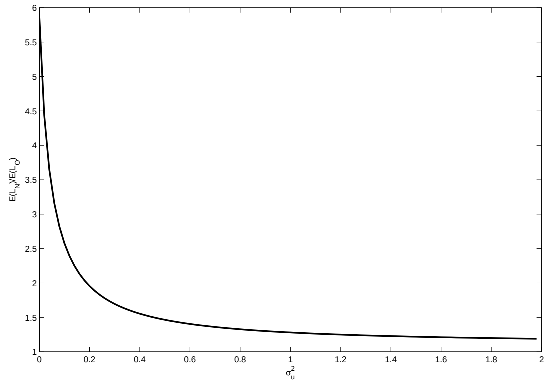


Figure 7: Welfare effect of transparency: impact of σ_u^2

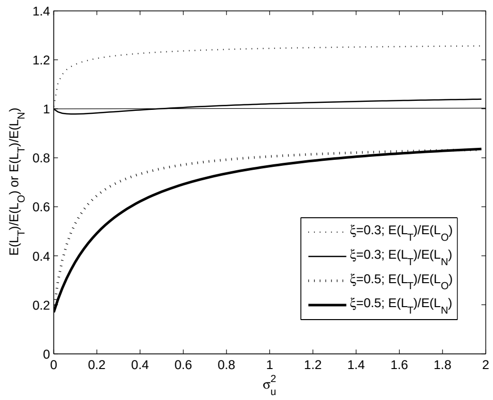


Figure 8: Comparison of relative losses

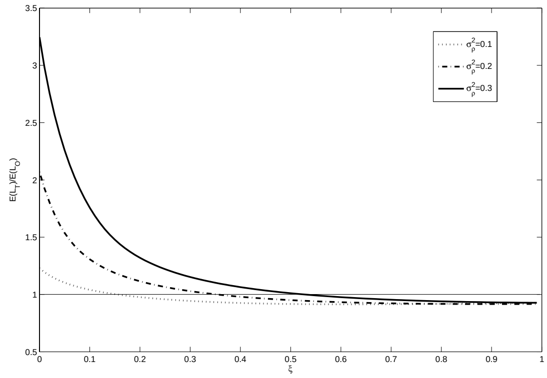


Figure 9: Welfare effect of transparency: impact of ξ

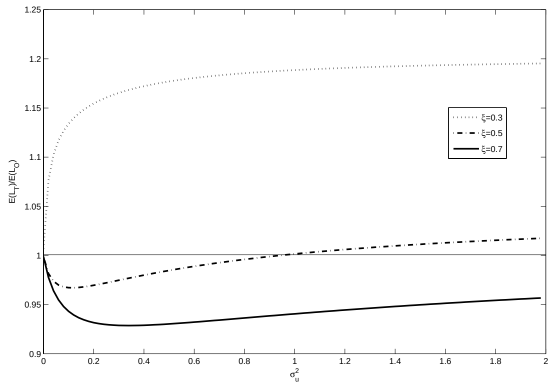


Figure 10: Welfare effect of transparency: impact of σ_u^2

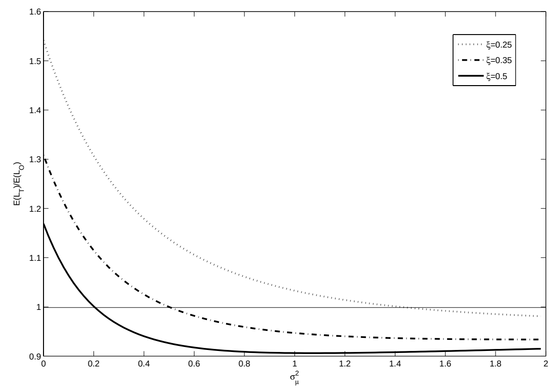


Figure 11: Welfare effect of transparency: impact of σ_μ^2

Crises and recovery in emerging markets: 'Phoenix miracles' or endogenous growth?

Marcus Miller and Lei Zhang

University of Warwick and CEPR — University of Warwick

June 2007

Prepared for 'Cycles, Contagion and Crises' Conference, FMG, LSE, June 2007

Abstract

In recent crises in emerging markets, currency devaluation and associated balance sheet effects have played a key role. Does this call for a new paradigm? After a comprehensive survey of many crisis episodes - including the US Great Depression - Calvo and colleagues from IDB offer the paradigm of a 'Phoenix Miracle'. It is a static, supply-side account where factor productivity falls in recession and rises promptly thereafter.

By contrast, we argue here that the real miracle in East Asia is economic growth. This can be rudely interrupted by external shocks: but then it restarts at a lower level of trend GDP. We show this in a simple model of endogenous growth, where investment is disturbed by balance sheet effects. Heterogeneity of production implies that both supply and demand effects are relevant as in the New-Keynesian paradigm for monetary policy proposed by Stiglitz and Greenwald, where monetary contraction affects both supply (of traded goods) and demand (for non-traded goods); and in the dynamic general equilibrium approach of Cook and Devereux (2006).

For discussion and suggestions, we thank Guillermo Calvo, Abhijit Banerjee, Neil Rankin and Ernesto Talvi and participants at the LSE conference; but responsibility for views expressed is our own. The authors gratefully acknowledge the research assistance of Javier Garcia-Fronti and Parul Gupta and the financial support of the ESRC, under Grant No. RES-051-27-0125 for Marcus Miller, and for Lei Zhang under Grant No. RES-156-25-0032.

Introduction

Financial crises in SE Asia sharply interrupted stellar economic progress in the region: for a year economic growth went into reverse. While it did not take long for growth to resume, output has followed a lower trend path. What happened, and why?

That these events were triggered by external capital market shocks is now widely accepted; in the terminology of Calvo and his colleagues at IADB, events in SE Asia followed a Systemic Sudden Stop in capital flows[1]. Liability dollarisation and the balance sheet effects that accompanied currency devaluation played a key role in depressing investment and supply, as did the high interest rates used to defend the currencies. There is, it appears, agreement on the importance of a type Fisher effect[2] - a central element of so-called third generation models of crisis.

But how to explain the sharp falls of output that were almost synchronous with the devaluation - and the prompt recovery thereafter? Here there are sharp differences of view. Stiglitz (2006) has argued that these events are explicable in terms of the New-Keynesian paradigm for monetary policy he has developed in conjunction with Greenwald; and he was one of the leading critics of IMF policy for the demand-reducing effects of the tight monetary and fiscal policy prescriptions imposed as a condition for financial support in East Asia. Cook and Devereux (2006) also claim that careful calibration of a sticky-price open economy model subject to a sharp rise in borrowing costs fits the data for SE Asian economies.

But after a comprehensive overview of crises in there and elsewhere, particularly in Latin America, Calvo and colleagues have arrived at a different conclusion; that the fall of output was attributable to adverse supply-side effects sharply lowering total factor productivity, whose prompt reversal they call a Phoenix miracle, CIT (2006). It is not a matter of remembering the Keynesian economics of depression: it is a matter of understanding how supply contracts in a financial crisis.

[1] Jeffrey Sachs said at the time that the SE Asia crisis was like a bank run.

[2] So-called by Calvo et al. (2006) after Irving Fisher who emphasized the role of liability effects in the US Great Depression.

To capture the setting of ongoing growth in SE Asia, we first consider these issues in the framework of a simple model of endogenous growth. In this context, the Balance Sheet effect certainly interrupts growth: but does not in itself cause a sharp fall in output. A temporary fall in TFP will cut output, of course: but so too does a temporary fall of output below existing capacity. The endogenous growth model used is of a closed economy, but the arguments go through in a model of a small open economy with a Balance sheet effect, as we indicate in the following section using a popular third-generation model of Aghion, Bacchetta and Banerjee.

Shocks to demand and supply may better be distinguished when account is taken of the distinction between traded and non-traded goods. Supply side effects will be more relevant for traded goods and demand effects for non-traded, as in New-Keynesian paradigm for monetary policy proposed by Stiglitz and Greenwald (where devaluation and tight money restrict aggregate demand and disrupt supply in the traded goods sector); and in the dynamic, sticky-price model of a small open economy of Cook and Devereux (2007).

In a profound financial crisis, the collapse of banks may well lead to economy-wide loss of efficiency, and we sketch how this can co-exist with the collapse of demand. With sticky prices and flexible wages, the perspectives of Calvo and the New-Keynesian may not be mutually exclusive.

1. Literature review

The deterioration of corporate balance sheets is a key element of recent crises in emerging markets, as emphasized in so-called 'third generation' models. But the role of debt in triggering economic contraction has a much longer pedigree: what Irving Fisher referred to as 'debt deflation' operating in the US Great Depression provides a historical precedent. But for Fisher, balance sheets deteriorated as falling prices raised the real value of nominal debt[3]: while in emerging markets it is devaluation against the dollar that raises the local currency value of debt contracts denominated in foreign

3 Note that, when the US left the gold standard, President F.D. Roosevelt cancelled the Gold Clause in public and private debt.

currency[4]. In the Biblical terminology of Eichengreen and Hausmann, the impact of the Fisher effect on private enterprise in emerging markets is a legacy of Original Sin. In Latin America, for example, the 'fear of floating' described by Calvo and Reinhart (2002) may be attributed to the balance sheet effects of devaluation and to the high interest rates that may be needed to check a fall in the currency.

Keynesian models can be adapted to take account of the Fisher effect. In his analytical afterthoughts on the Asian crises, for example, Krugman (1999) used a *demand-side* account of a small open economy to argue that "a loss of confidence by foreign investors can be self-justifying, because capital flight leads to a plunge in the currency, and the balance-sheet effects of this plunge lead to a collapse in domestic investment." In a much more detailed framework, Céspedes et al. (2003) show that for a highly dollarised open economy, the asset price effects of devaluation can overwhelm trade effects, leading to a contraction of aggregate demand.

In the "third generation" approach developed by Aghion, Bacchetta and Banerjee (2000), however, it is *supply-side* effects that play a central role. In the first place, the trigger for crisis is an unexpected idiosyncratic, permanent fall in productivity which lowers expected future supply. In the absence of a corresponding contraction of future money, the currency is expected to be weaker in future. Anticipation leads to prompt and unexpected devaluation with adverse balance sheet effects on investment further reduces future supply. Current output is unaffected by the crisis, however.

Productivity effects also play a key role in account of Calvo et al (2006), hereafter CIT, though here they are endogenous and temporary. The causal factor is a Sudden Stop which leads to devaluation and a fall in productivity of currently installed plant and equipment. "Sharp nominal (and real) currency devaluation in the presence of Liability Dollarisation may have worked in Emerging Markets as a new version of Fisher's Debt Deflation syndrome, and may be central in explaining output collapses." CIT (2006 pp.10, 11). How this might occur is explained by a partial equilibrium model, where all output can be sold, but a sharp rise in the ex ante short-term cost of borrowing not only reduces inventories and but also induces the firm to

4 As a consequence, debt can increase in real terms even when domestic prices rise.

sell physical capital[5] to finance inventories: so firm output falls. (Alternative accounts might draw on models of equilibrium shift: e.g. Diamond (1982), Diamond and Dybvig (1983) and Allen and Gale (2007).) Happily these effects are temporary and productivity soon recovers without recourse to outside finance for the firm: this is the Phoenix miracle.

Stiglitz and Greenwald, by contrast, stress both the demand-side and supply-side effects of tight monetary policy at the time of the crisis. In the case of Korea, for example, where interest rates were pushed above 25% to try to stabilise the currency, they argue that:

> it was exporters' failure to respond to the huge exchange rate reduction – which should have stimulated demand - which makes it clear that the economy was not just responding to a fall in aggregate demand. Such consequences were inevitable, unless the producer could obtain cheaper credit elsewhere … or unless wage and price adjustments were sufficiently large to compensate for the huge increase in capital costs. In practice, even reductions in real wages of 20% or more did not suffice. Thus monetary policy had the usual effect on aggregate demand (amplified by the adverse effect of increased bankruptcy probabilities on firm demand) … but they also had huge effects on aggregate supply.

(Towards a New Paradigm in Monetary Economics, p.264)

In a detailed exercise in calibration, Cook and Devereux find that they can account for the macroeconomic data on prices and output in South Korea, Malaysia and Thailand using a two-sector dynamic model with sticky prices to assess the response to an unanticipated rise in the country risk premium. They summarize the qualitative effects as follows:

> As all three countries are net debtors, the interest rate rise has a negative income effect, reducing optimal consumption…. [Given that the shock persists] the interest parity condition will imply an immediate depreciation of the nominal exchange rate. …In equilibrium, there is a decline in demand for domestic goods. The exchange rate depreciation does not immediately increase the quantity of exports, because exporters practice local currency pricing
>
> The impact of the risk premium shock on production depends on the stance of monetary policy. …The contractionary monetary policy, observed in the data and matched by both monetary rules in our model, involves an increase in domestic rates. This mitigates the immediate real exchange rate depreciation, and hence leads to a greater decline in absorption and output in domestic traded and non-traded goods. … we anticipate a bigger fall in output in the non-traded sector.

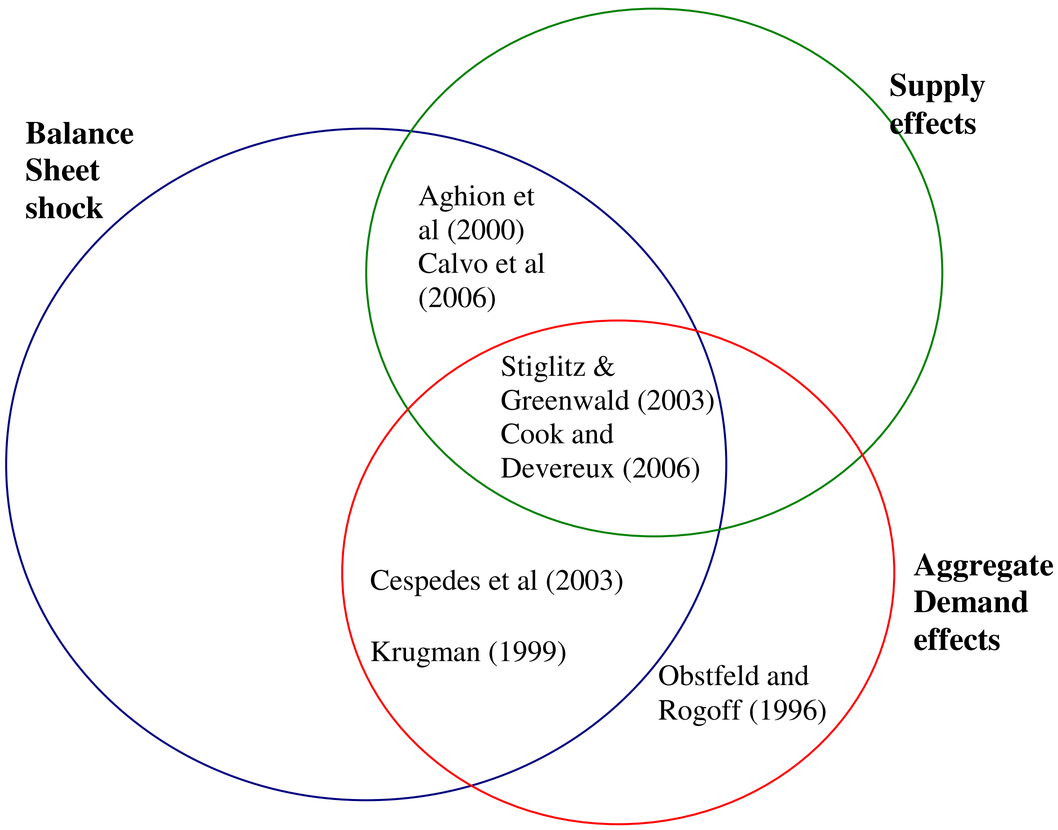
'Accounting for the East Asian Crisis'p.p.737,8

---

[5] Capital is apparently perfectly liquid as it can be sold at a fixed price with no transactions cost.

Views on the role of financial effects and the transmission mechanism are summarised in the following Venn diagram. The papers just discussed appear in the left circle as they ascribe a key causal role to the financial shocks – sharp increases in net liabilities amplified by high interest rates used to defend the currency; but they differ on the transmission mechanism -- whether they work through demand or supply.

Figure 1 Financial shock, demand and supply



An interesting contrast is provided by the treatment of the small open economy in Obstfeld and Rogoff (1996, Chapter 10.2), where the output of non-traded goods, whose prices are fixed in the short run, varies with the exchange rate. Monetary expansion leading to unanticipated devaluation has an unequivocally positive impact on demand and production of these goods as they become cheaper than flex-price traded goods (see Appendix). Maybe because it was written before the Asian crises, there is no role for the balance sheet effects in this model.[6]

[6] Note that the eclectic approach of Cespedes et al. (2003) can deliver positive or negative effects of devaluation on output, depending on the relative importance of balance sheet considerations.

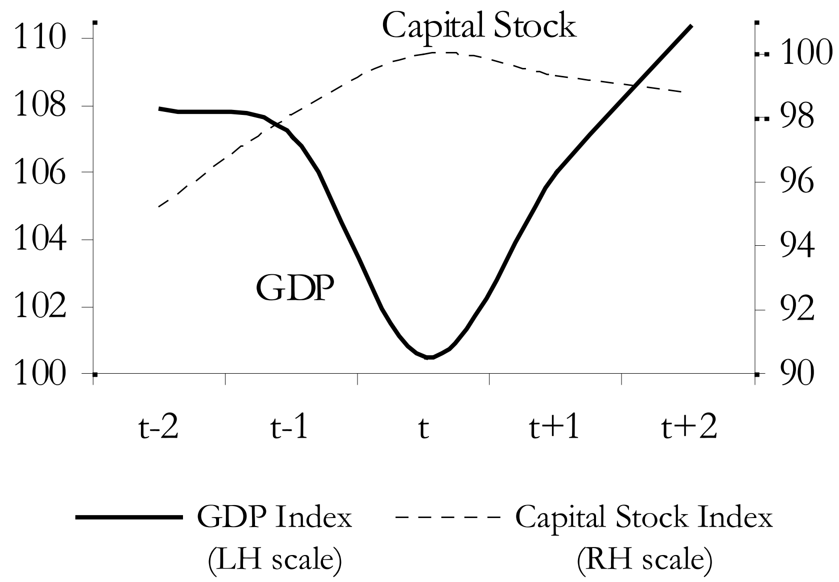
In the monograph on *Understanding Financial Crises*, Allen and Gale (2007) draw attention to the critical role of bank intermediation --- and its fragility in the face of real shocks to investment returns. (Shocks to net returns may well arise from the adverse balance sheet effect of currency devaluation, as discussed in Moheeput, 2007.) One criterion for assessing the models discussed below, indeed, might be how well they reflect the various aspects of financial crisis described by Allen and Gale.

In the next section we show the stylised facts that Calvo and colleagues events describe as a Phoenix miracle. That their account takes no account of ongoing economic growth in the region is indicated briefly where the U shaped curve of CIT is shown alongside with the long run growth. But the data from India provide a striking contrast.

II. Phoenix miracle; or interrupted growth?

In CIT (2006), the stylized facts characterizing 3S output collapses are presented in a series of fascinating graphs averaging data across affected countries. To illustrate, Figure 2 reproduces the typical path of GDP so derived; and a matching index of the capital stock.

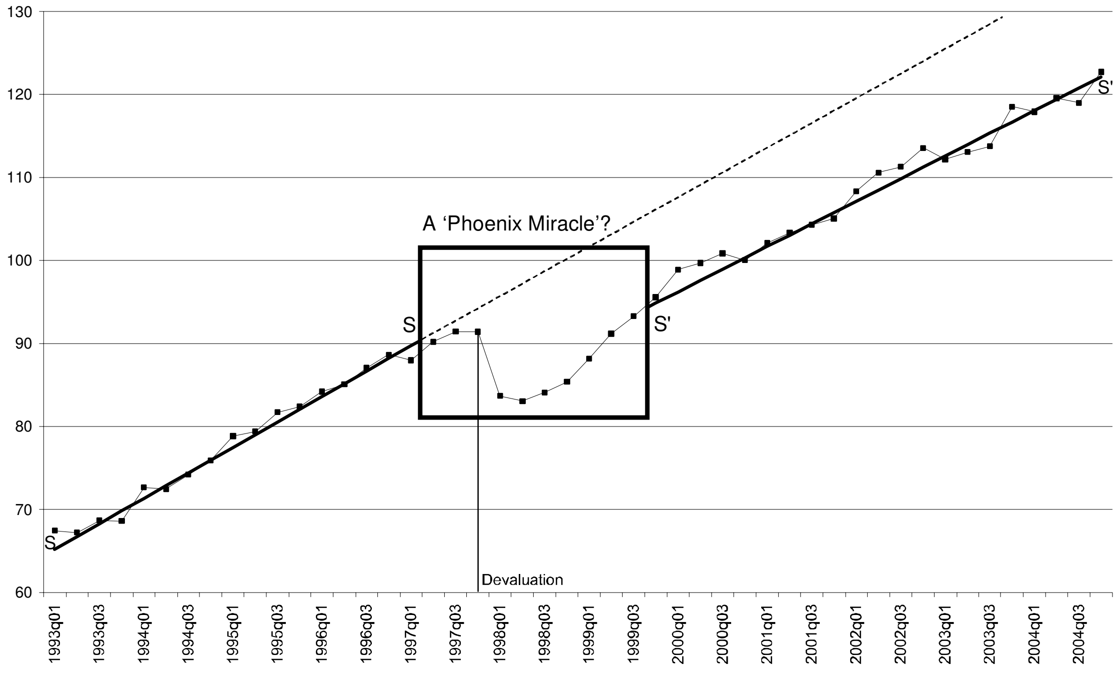
Figure 2 'Phoenix miracle'



“These episodes”, they observe, “are characterized by two salient features. First, there is a dramatic collapse in output (…the average fall in GDP is 10 percent) accompanied by a collapse in credit, but without any correspondingly sharp collapse in either physical capital or the labour force. Second, recovery to pre-crisis output is swift and “credit-less”… *Thus, although a credit crunch appears to be central for explaining output collapse, recovery can take place without credit*. This remarkable phenomenon that resembles the feat of the proverbial bird “rising from its ashes” prompted us to call it Phoenix Miracle.”

A longer run perspective may be obtained by plotting GDP for some time before and after the crisis. The longer run of data allows one to fit two trend paths, one before the crisis and one after (with quarterly seasonal adjustment around the split trend), so as to provide a crude estimate of potential GDP, of long run supply[7].

Figure 3 Korean GDP



Source: IMF:IFS. (Note: output volume is normalised to 100 in 2000; and the data seasonally adjusted by authors)

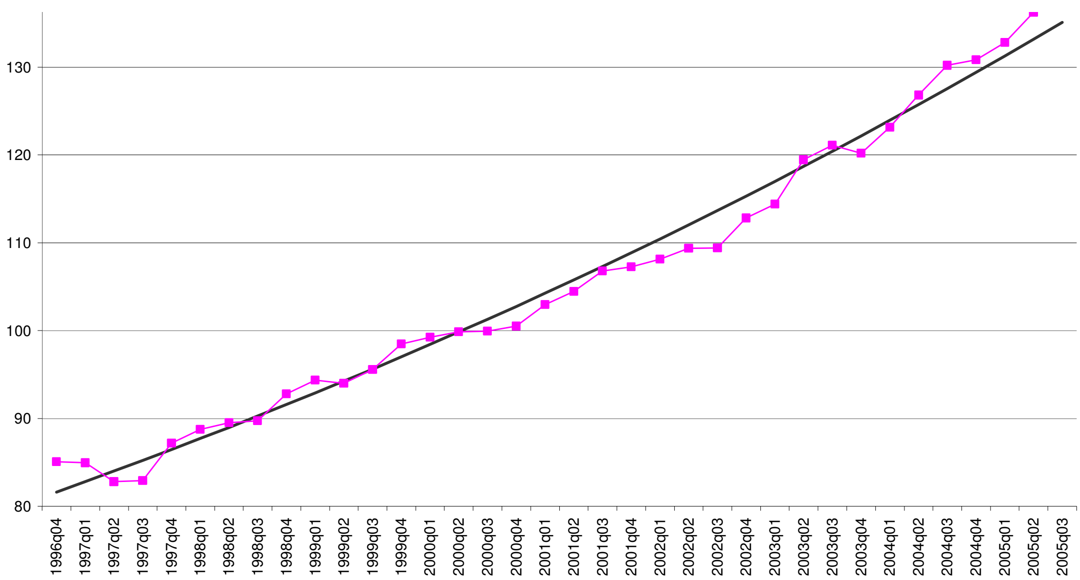
For the case of Korea shown in Figure 3, starting in the box showing data around the time of the SE Asia crisis, one sees that GDP follows broadly the same trajectory shown in the earlier figure. Looking outside the box, however, gives a new

[7] Aghion and Banerjee (2005) provide a similar graph for Indonesia.

perspective: though the trends fitted before and after have much the same slope, there is a difference of about 10 percent in the level. From a growth perspective, therefore, there appears to be a permanent loss of potential output. (Figures for Thailand and Malaysia show a similar pattern of a down-shift of trend potential. [8])

In addition to the sharp fall in output summarised in Figure 3, there also appears to be a down-shift in the trend path of output growth. As to why this might be so, the behaviour of capital stock shown in Figure 3 provides a clue: on average net capital formation effectively ceases during the crisis.[9] In the next section we sketch an endogenous growth model where a temporary dip in GDP leaves a permanent mark on potential GDP.

Figure 4 Indian GDP



Source: IMF:IFS. (Note: output volume is normalised to 100 in 2000; and the data seasonally adjusted by authors)

---

[8] Note also that, in their analysis of the earlier Tequila crisis, Kehoe and Ruhl (2007) find that they have to augment their supply-side model with an exogenous unanticipated but permanent fall in TFP of about 6% in order to make it fit the Mexican data.

[9] The lack of external corporate finance in recovery is part of what CIT have described as a miracle. The behaviour of real wages must surely be taken into account in this connexion, as a shift in factor shares in favour of profits will increase the potential supply of internal funds for financing recovery. While output falls by about 10% on average, CIT report that *real wages fall by about a quarter* in emerging market crises: and from this there is, apparently, no recovery.

The data for output in India shown in Figure 4 provide a striking contrast, with very little evidence of the shock that shook its eastern neighbours. Indian capital markets were kept relatively closed during the 1990s and the country escaped the currency and output shocks suffered by many of its neighbours, Williamson (1999). Figure 4 shows little sign of any shift of potential in India at the time of the SE Asian crisis, as a single exponential trend seems to fit the data.

III. Supply, Demand and Crises: an AK approach

As Fischer (2001) has emphasized, financial crises in South-East Asia involved an initial reversal of the growth rate, followed by prompt recovery: so output traces a V-shape and as shown in Figure 2. To capture the setting of ongoing expansion, we explain these output effects in stylised fashion using a simple AK model of endogenous growth which incorporates balance sheet effects and their impact on productivity and/or aggregate demand.

Let potential supply (Q) be determined by the capital stock (K) so

$$Q = AK \tag{1}$$

and net capital formation be defined as

$$\Delta K = I - \delta K \tag{2}$$

where Δ is the forward difference operator, and
δ is the rate of exponential depreciation.

Gross investment depends on the flow of savings and the impact of balance sheet effects, so

$$I = sQ - \beta K \tag{3}$$

where s is the propensity to save and βK is an adverse balance sheet effect, a Fisher effect, see Allen and Gale for a thorough treatment of the effects of financial crisis.

In the absence of balance sheet effects, one obtains the canonical growth rate, namely:

$$\frac{\Delta Q}{Q} = \frac{\Delta K}{K} = \underbrace{sA - \delta}_{\substack{\text{Canonical} \\ \text{endogenous} \\ \text{growth} \\ \text{rate}}} \equiv g_c \tag{4}$$

Phoenix Miracle

If an *adverse balance sheet effect* were simply to cut investment by βK for one period, the growth of capital and output following the shock will simply fall by β, so

$$g = \frac{\Delta Q}{Q} = \frac{\Delta K}{K} = sA - \delta - \beta = g_c - \beta; \tag{5}$$

but there will be no fall in output at the time of the shock. Adding a *temporary fall in productivity* at the time of the shock[10], so $Q = (1-\gamma)AK$, will reduce savings so capital will grow more slowly after the shock, specifically

$$\begin{aligned} \frac{\Delta K}{K} &= s(1-\gamma)A - \delta - \beta = g_c - \beta - \gamma sA \\ &= g_c - \beta - \gamma(g_c + \delta) \end{aligned} \tag{6}$$

Further, the fall of productivity means that output growth going into the recession will be approximately

$$\frac{\Delta Q}{Q} = sA - \delta - \gamma = g_c - \gamma$$

And there will be a sharp recovery coming out, namely

$$\frac{\Delta Q}{Q} = s(1-\gamma)A - \delta - \beta + \gamma A = g_c - \beta - s\gamma A + \gamma.$$

Consider for example the case where the balance sheet effect just offsets canonical growth, i.e. $\beta = -(sA - \delta) = -g_c$, but the productivity effect is twice as large, i.e. $\gamma = -2g_c$. In this case output growth will fall to $-g_c$ on entering into recession but recovery will take place at almost twice the canonical rate. This is illustrated in Figure 5.

[10] Could one appeal to Allen and Gale's analysis to justify this? Or Diamond (1982)?

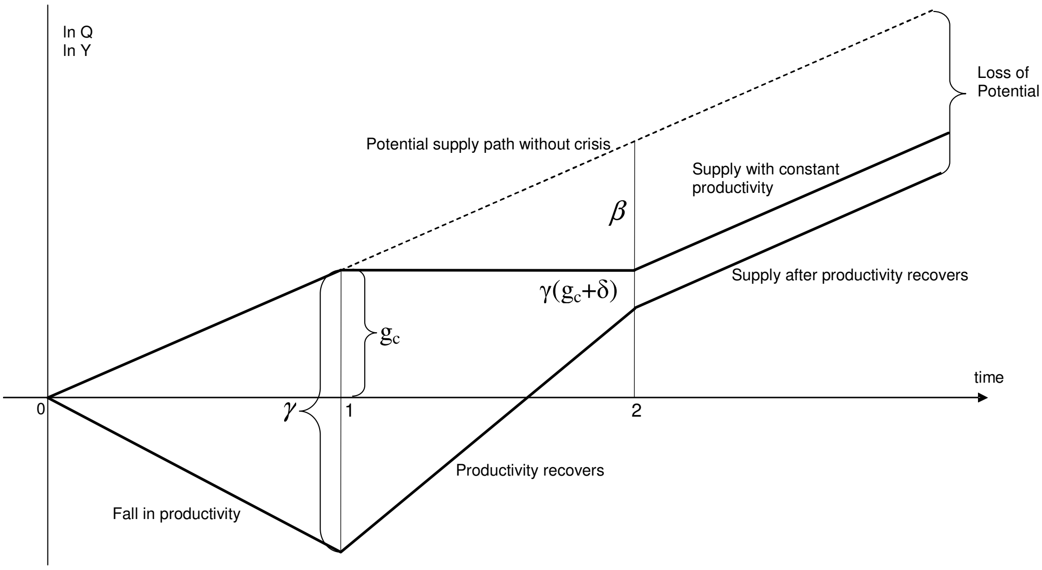


Figure 5 Financial shock: short and long run effects on supply.

The impact of an adverse balance-sheet effect induced by currency devaluation at period 1, *assuming no productivity shock,* means that output is not affected in period 1 but falls below the pre-existing trend by β% in all subsequent periods, as shown by the upper solid line in the figure[11]. If the balance sheet effect triggers a fall in TFP so that the growth rate of GDP changes sign (as was roughly the case for countries in SE Asia), then the dip will lower potential supply yet further by the amount, $\gamma(g_c+\delta)$ even if productivity recovers promptly in period 2. With the productivity recession, the lower bold line shows output exhibiting the familiar V-shape in the period of crisis and recovering promptly thereafter (but to a lower trend) in period 2. This satisfies the output pattern of a Phoenix miracle and the longer term downshift of supply shown in Figure 2.

A similar pattern may be observed in recent crises affecting countries in Latin-America, as Talvi (2006) indicates. Using Central-American GDP as proxy for the trend of potential supply, he finds the characteristic V-shaped recession. Economic recovery, accompanied by the redistribution of income in favour of profits but without external credit, is fairly rapid; but it does not take output back to the previous trend.

[11] Drawn on the convenient, but not essential, assumption that the balance sheet effect is sufficient to wipe out the effects of one year's growth, i.e. $\beta = sA - \delta$.

Keynesian Recession

Much the same results follow if recession causes demand to fall below supply. Consider a simple Keynesian-multiplier account of the same phenomenon, assuming, as before, an adverse balance sheet effect that cuts investment by βK for one period. If this fall of investment has a multiplier effect on income, then there will be a recession as output falls beneath the capacity by βK/s. These deviations from trend growth will be temporary if investment and demand recover promptly as balance sheet problems are resolved, but output will lie below its previous trend.

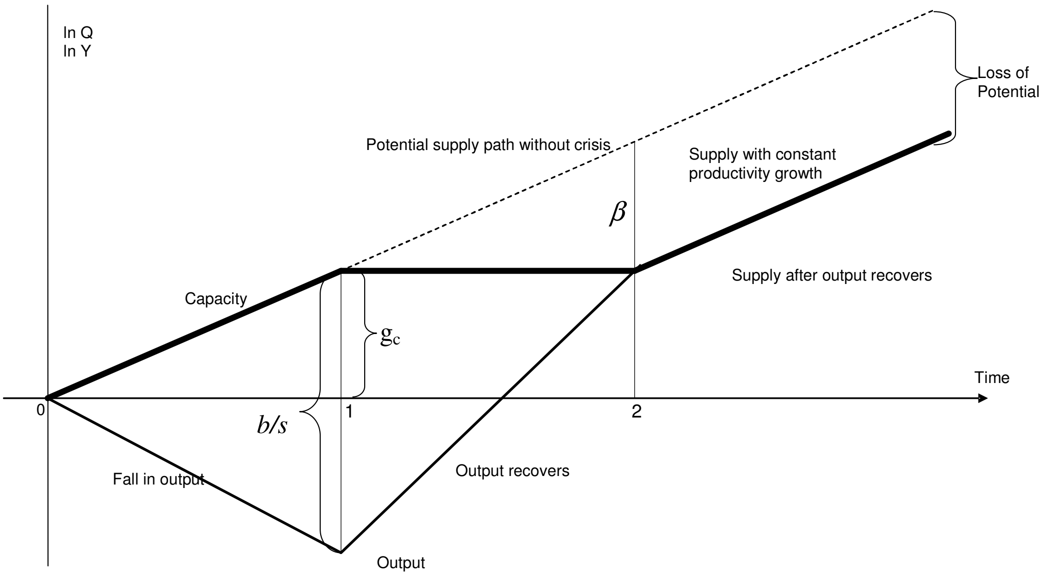


Figure 6 Financial shock: effects on demand and supply

In this case, capacity will continue to grow as $\frac{\Delta Q}{Q} = sA - \delta = g_c$ in the period of crisis but output evolves as $\frac{\Delta Y}{Y} = g_c - \frac{\beta}{s}$. The temporary fall in investment will lower future potential, but there will be a sharp recovery coming out of recession, as shown in Figure 6.

IV. Financial shocks to a small open economy

The endogenous growth model just considered was for a closed economy. How to take account of open economy aspects? The comprehensive exercise conducted by Cook and Devereux (2006) involves a sophisticated dynamic general equilibrium model of a small open economy; and requires calibrating a set of more than thirty equations. As an alternative approach, we make use of a popular supply-side model of monetary policy in a small open economy written to capture to the exchange rate and balance sheet effects operating in East Asia by Aghion, Bacchetta and Banerjee (2000), hereafter ABB. It was, we gather, designed as a workhorse to analysing some of the issues involved.

Specifically, we use this framework to consider how financial shocks might have both supply side and demand effects, leaving longer run growth aspects to one side. In an open economy financial restriction could, for example, affect the supply of traded goods (for which there is elastic demand), and the output of non-traded goods (where it is only local demand that matters). Two changes inspired by the work of Calvo and his colleagues at IDB are introduced into the ABB framework. First *the cause of the crisis is taken to be an external financial shock* (not an exogenous shock to domestic TFP). Second *financial conditions are assumed to impede the expansion of exports* in the short run so that the economy lacks the stabilising feature of unlimited foreign demand that assures full employment of resources even in the short run. The results can be interpreted as the effect of a Sudden Stop in capital flows in an economy temporally deprived of an automatic stabiliser.

The basic ABB model in summary form

The ABB model is a dynamic supply-side model which focuses on the balance sheet effects of devaluation on the private sector in a small open economy. With liability dollarisation and one-period of price stickiness for the traded good, a rise in the price of the dollar generates adverse balance sheet effects; so investment is cut back, reducing productive potential in the next period. This "third generation" account offers a persuasive channel for the transmission of the exchange rate effects to the supply side; and the multiplicity of equilibria opens up the prospects of sudden shifts in the exchange rate. Key features of this widely cited two-period model may be summarised as follows.[12]

There is full capital mobility and uncovered interest parity holds. Purchasing Power Parity (PPP) for traded goods also holds, except in period 1 when an unanticipated shock leads to a deviation as prices are preset, but other variables — the nominal exchange rate in particular — are free to adjust. The actual timing of events in period 1 is: the price of traded output is pre-set according to the ex-ante PPP condition and firms invest; then there is an unanticipated shock[13], followed by the adjustment of interest rate and the exchange rate; finally, output and profits are generated, with a fraction of retained earnings saved for investment in period 2. Together with investment funded by lending, this determines the level of production in the second period, when prices are flexible so PPP is restored.

An attractive feature of this model is that the equilibrium can be found as the intersection of two schedules relating the exchange rate and output in the following period, called the IPLM curve and the W curve. The former, as the name suggests, is a combination of the Uncovered Interest Parity, money market equilibrium and the PPP condition for the second period. Formally, it is written as:

$$E_1 = \frac{1+i^*}{1+i_1} \frac{M_2^s}{L(Y_2, \bar{i}_2)} \qquad (7)$$

where E_1 is the exchange rate for the first period, i^* is the foreign interest rate, i_1 i1 and $\bar{i}_2$ are domestic interest rates for periods 1 and 2, M_2^s and Y_2 are money supply and output in period 2, and $L(Y_2, \bar{i}_2)$ is the money demand function. This IPLM curve is downward sloping in the E_1 and Y_2 space because higher output in the second period increases money demand (i.e., higher L(.) given interest rate in period 2) and so strengthens the exchange rate (note that M_2^s is given).

[12] The relevant equations are given in Appendix B.

[13] If the shock is anticipated, the expected price adjustment eliminates the balance sheet effect, Becker (2006).

The second of the two schedules, the W-curve, characterizes the supply of output on the assumption that entrepreneurs are credit-constrained. (The production function is assumed to be linear in capital stock, which *depreciates completely at the end of the period*.) Total investment consists of last-period retained earnings together with borrowing (in both domestic and foreign currencies, with proportions given exogenously) which is limited to a given fraction $\mu_t(i_{t-1})$ of retained earnings. The introduction of $\mu_t(i_{t-1})$ (with $\mu_t' < 0$) captures credit market imperfection. The W-curve is specifically given by

$$Y_2 = \sigma(1+\mu(i_1))(1-\alpha)\left[Y_1 - (1+r_0)D^c - (1+i^*)\frac{E_1}{P_1}(D_1 - D^c)\right] \quad (8)$$

where σ is the productivity parameter, α is the fraction of output consumed in each period, D_1 is the total level of borrowing and D^c is its domestic currency component. The W-curve so constructed is a downward sloping straight line in E_1 and Y_2 space[14]. Clearly this formulation captures the contractionary effect of devaluation on the supply-side, i.e. the Fisher effect.

ABB (2000) use the framework to analyse the policy dilemma posed by a negative TFP shock which lowers anticipated future output. With no changes in anticipated future money supply, the expected higher future price level and lower value of the currency induces current devaluation, which triggers adverse balance sheet effects and damages investment[15]. Raising interest rates to strengthen the currency risks further contraction of investment and in future output.

Supply constrained exports and aggregate demand failure

The ABB analysis assumes that there is no aggregate demand failure resulting from the productivity shock: investment may fall due to balance sheet effects, but this does not affect output at the time of the currency crisis. The demand for exports of a small open economy is typically assumed to be unlimited: so exports, in theory, can adjust

[14] Note that Y_2 is set to zero if the right hand side of (2) turns out to be negative, where $Y_2 = 0$ signifies the depression level of output.
[15] Note that the alternative monetary policy options discussed by Cook and Devereux (inflation and exchange rate targeting) might well avoid these adverse effects.

to provide an 'automatic stabiliser' for demand shocks to demand. The data, however, are not consistent with this reassuring hypothesis. Calvo and Reinhart (2000), for example, find that in case of an emerging market currency crisis, exports typically *fall* before recovering to their pre-crisis levels: the lag before recovery is 8 months or, with a banking crisis, 20 months. In an investigation of devaluations in emerging economies, Frankel (2005, p.157) concludes "that devaluation is contractionary, at least in the first year, and perhaps in the second as well."

To account for the 'Phoenix Miracle', CIT (2006) have, as we have seen, invoked a large (but temporary) economy-wide productivity shock affecting current output: specifically they argue that there is an immediate contraction of supply across all sectors of the economy, attributable in large part to the reduction of working capital[16]. For Stiglitz and Greenwald, such supply-side factors certainly apply to exports -- but demand effects operate elsewhere.

Following CIT (2006), let us postulate a temporary fall in productivity across the economy, captured by a fall in an 'efficiency term' η appearing in the supply function in the period that the crisis breaks. This is assumed to determine the behaviour of exports relative to their trend growth rate- but not the fall in total output as in the Phoenix Miracle account as aggregate output is demand-determined due to the multiplier effect of depressed investment on GDP. (In support of the Keynesian specification of demand determination used here, note that firms are in any case *credit-constrained* in the ABB model; and where devaluation is accompanied by a banking crisis ("twin crises"), both consumers and producers will typically be denied access to new credit.)

With demand-determined output, the fall of investment will cut current output and consumption. Specifically, let output in period 1 be determined as follows:

$$Y_1^D = A_0 + \gamma\beta\left(Y_1 - D_1^*\right) + \left(1+\mu_2\right)\left(1-\beta\right)\left(Y_1 - D_1^*\right) + \bar{X} - mY_1 \qquad (9)$$

[16] In their partial equilibrium analysis, this is augmented in deep crises by the sales of physical capital to further economise on inventories: what happens in general equilibrium is not clarified, however.

where $D_1^*(E_1) = (1+r_0)D^c + (1+i^*)(E_1/P_1)(D_1 - D^c)$ is the total cost of debt service and Y_1 is aggregate demand measured in constant prices. The first term, A_0, represents autonomous expenditure (which is not related to debt or current income). The second term indicates how consumption demand depends on income and debt, where $\beta < 1$ is the labour share of income and $\gamma < 1$ is the fraction spent on consumption. The third term is demand for investment with $(Y_1 - D_1^*)$ representing corporate profits net of borrowing costs, and μ the credit multiplier. The last two terms represent net exports: while imports vary proportionally with current income, the export volumes in the current period are taken to be a fraction ϕ of aggregate supply $\bar{X} = \phi\eta Y_1^S$
where Y_1^S is the supply in the absence of the shock and the efficiency parameter η is given by

$$\eta = \eta(\mu_2(i_1) - \mu^*) = \begin{cases} 1 & if \quad \mu_2(i_1) - \mu^* \geq 0 \\ <1 & if \quad \mu_2(i_1) - \mu^* < 0 \end{cases} \qquad (10)$$

When interest rates are relatively low, such that the credit multiplier is above the threshold μ^*, the efficiency parameter η is equal to 1. But when interest rates are sufficiently high to cause the credit multiplier to fall below the threshold μ^*, the efficiency parameter falls below 1. (CIT (2006) go further, arguing that supply will make a jump decrease as interest rates go pass the threshold that leads to a deep crisis.)

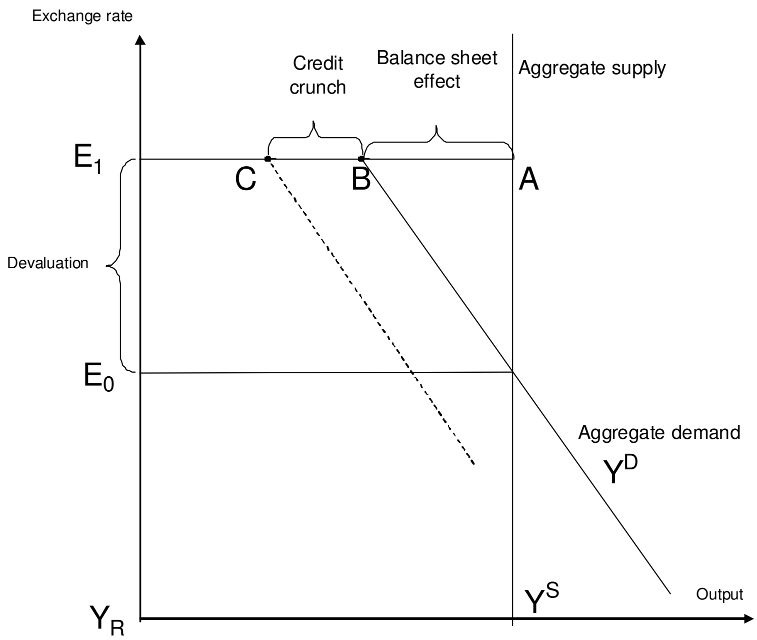
The failure of export volumes to stabilise demand means that a collapse of investment (due to balance sheet effects, for example) can reduce realized output in the current period (as well as supply potential in the next period), as can be seen from the solution for current output:

$$Y_1^D = \frac{A_0 + \bar{X} - D_1^*(E_1)\left[\gamma\beta + (1+\mu_2(i_1))(1-\beta)\right]}{1+m-\left[\gamma\beta + (1+\mu_2(i_1))(1-\beta)\right]} = \frac{A_0 + \phi\eta Y_1^S - \xi D_1^*}{1-\xi+m} < Y_1^S \qquad (11)$$

where $\xi = \gamma\beta + (1+\mu_{t+1})(1-\beta)$ and $1 > 1 - \xi + m > 0$. The predetermined factors in the numerator include debt service and exports volumes; and the term $1/(1-\xi+m)$ is a Keynesian-style open economy multiplier, where 1-ξ is the marginal propensity to save and m is the marginal propensity to import.

How demand failure can lead to prompt contractionary devaluation is indicated in Figure 7, with output in period 1 on the horizontal axis and the exchange rate in period 1 on the vertical. As it depends essentially on output and interest rate in the previous period, aggregate supply (in the absence of the crisis effect on TFP) appears as a vertical line Y_1^S. Aggregate demand, however, moves inversely with the current exchange rate due to the adverse balance-sheet effects of a devaluation which raises the price of a dollar from E_0 to E_1 and increases $D_1^*(E_1)$ in equation (2). At E_1, for example, demand has fallen by AB.

Figure 7 Aggregate demand and supply in period 1



In a "*twin crisis*", where devaluation is accompanied by a credit crunch, aggregate demand will fall even further.[17] Two scenarios are considered here. The first is that a devaluation results a mild credit crunch where credit multiplier $\mu_2(i_1)$ falls on impact but remains above the threshold $\mu*$. In this case, the aggregate supply in period 1 is

[17] Becker (2006) discusses the conditions under which a credit crunch will reduce output in this context.

unaffected at the level of Y_1^S, and aggregate demand shifts further to the left due to the credit crunch (as shown by the dotted line in the Figure). At E_1, further demand fall due to the credit crunch is measured by the distance BC in Figure 7. The second scenario, a severe credit crunch, which lowers $\mu_2(i_1)$ below the threshold value μ^*, will have additional supply side effects. First, the severe credit crunch shifts current period aggregate supply from Y_1^S to $\eta Y_1^S < Y_1^S$. This fall in supply reduces exports, causing additional contraction in aggregate demand. So output could fall below C at E_1. As for the effects of raising interest rates to defend the exchange rate, they are demand-contractionary: an increase in the period 1 interest rate will reduce Y_1^D as high interest rates impact adversely on the credit multiplier and so on investment.

In Table 1 we compare and contrast the standard ABB model, where output is supply-determined, with what occurs when exports are predetermined. For ABB, an adverse devaluation-induced shock to the balance sheet in period 1 has no effect on period 1 output (which is determined by previous period investment), but cuts it in period 2 via reduced capital accumulation, see column 1.

Table 1: How demand failure modifies output levels. [18]

	ABB(2000)	As modified
Y_1	$Y_1^S(ABB) = \sigma\left[1+\mu_1(i_0)\right](1-\alpha)\left[Y_0 - D_0^*\right]$ $Y_1 = Y_1^S(ABB) = Y_1^D$	$Y_1^D = \left[\gamma\beta + (1+\mu_{t+1})(1-\beta)\right](Y_t - D_t^*) +$ $+ A_0 + \bar{X} - mY_t$ $Y_1 = Y_1^D < Y_1^S = Y_1^S(ABB)$
Y_2	$Y_2^S(ABB) = \sigma\left[1+\mu_2(i_1)\right](1-\alpha)\left[Y_1 - D_1^*\right]$	$Y_2^D = Y_2^S < Y_2^S(ABB)$

When there is excess supply in period 1, however, the impact of an anticipated currency collapse is more immediate and more damaging. Balance-sheet effects reduce investment in period 1 directly: but this triggers a contraction of income within

[18] Note that in table 1, we have followed ABB in assuming that output in period 2 is supply-determined. This does not mean that output in period 2 matches that of the ABB model, however: the contraction is greater because of the reduced investment associated with the fall in aggregate demand in period 1.

the period, which in turn leads to even less investment as profits fall. The knock-on effect on period 2 means that future supply is less than predicted by the ABB model.[19]

Adding endogenous growth

The simplifying assumption made by ABB that capital depreciates completely within one period highlights the effect of reduced investment in dramatic fashion: this period's investment is next period's capital stock! But the exaggerated rate of depreciation effectively rules out the growth-creating effects of capital formation. A more attractive alternative is a growth model (with depreciation well below unity), where lower investment cuts future supply and also leads to a sharp recession via its effects on aggregate demand. How can this be incorporated?

In the ABB model, output is defined by

$$Y_2 = \sigma\left(1+\mu\left(i_1\right)\right)K_2$$

and implicitly investment by $I_1 = (1-\alpha)\left(Y_1^D - \left(1+r_0\right)D^c - \left(1+i^*\right)\frac{E_1}{P_1}\left(D_1 - D^c\right)\right)$

(Note that with 100% annual depreciation $K_2 = I_1$).

To fit their analysis into an endogenous growth framework one could interpret (and modify) this investment equation along the lines of equation (5) in Section II, namely

$$\Delta K = S - \delta K = \sigma AK - \beta K - \delta K$$

i.e. to assume all saving is invested but to allow for only partial depreciation. Thus, with very little violence to their algebra, it appears that their model can be transformed into a model of endogenous growth.

V. Combining the New-Keynesian and Calvo accounts

In this section, the difference between supply-side view of CIT and the New-Keynesian account of Cook and Devereux (2006) is indicated schematically for closed economy before considering a compromise which may be appropriate in the case of severe financial crisis including bank closures.

[19] Cutting μ_1, credit multiplier corresponding to period 0, would have same effects on period 1 supply in both models.

In Figure 8, equilibrium prior to the financial shock is shown at point E where marginal product of labour (MPL) intersects the labour supply curve (labelled S). The supply-side approach of Calvo and colleagues is to reduce the marginal product of labour to MPL' (as indicated by the arrow pointing down from E); so equilibrium employment and real wage shift to the point labelled CIT.

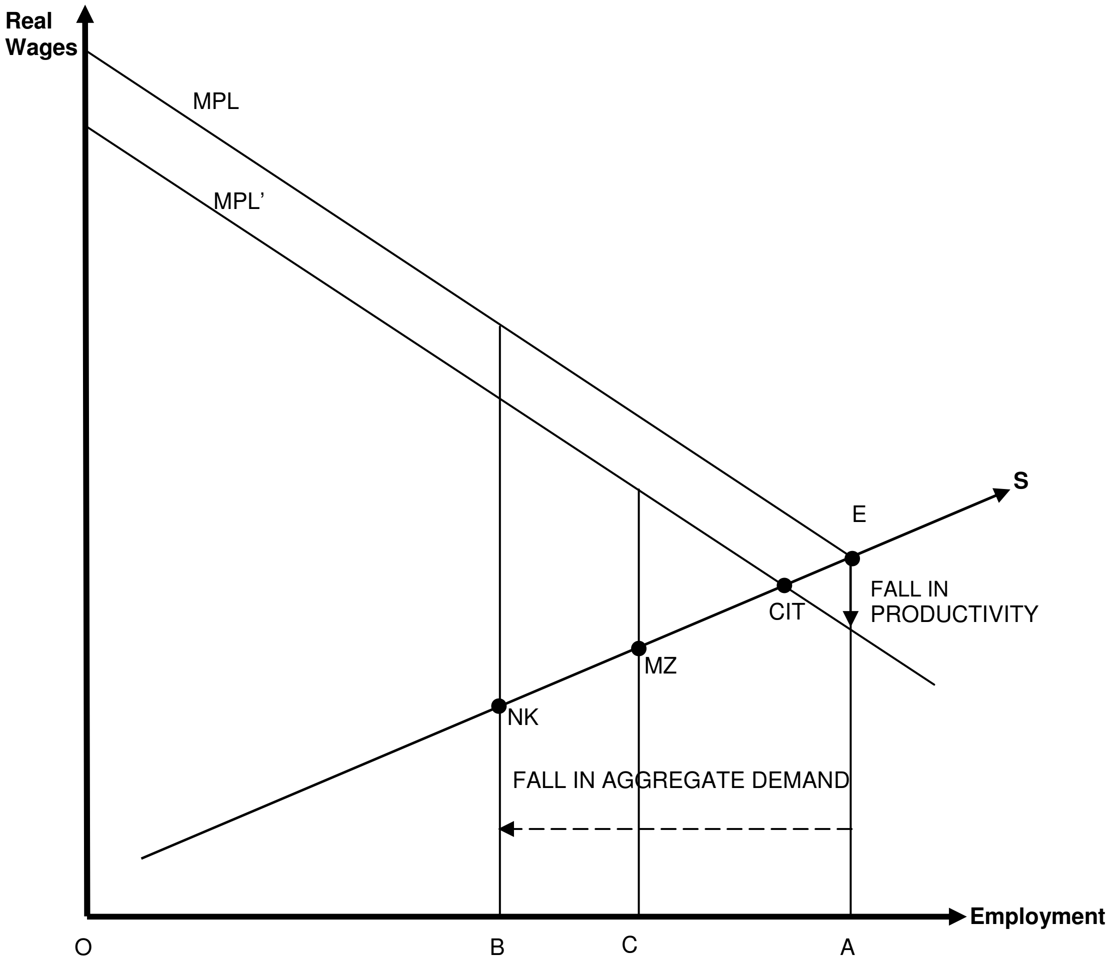


Figure 8 Combining a fall in the efficiency with the collapse in demand

By contrast, the New-Keynesian account focuses on the fall in aggregate demand, indicated the dotted line in the figure, driven perhaps by adverse balance sheets hitting investment. After an adverse demand shock, the level of employment can be found by integrating MPL from O to B (instead of from O to A), so the real wage falls as indicated by the equilibrium labelled NK. This interpretation, that demand for labour is obtained by inverting the production function and the real wage comes from labour supply, is based on Woodford (2003, Chap 3.4.2). It seems consistent with Cook and Devereux (2006) who assume sticky prices and flexible wages (although no figures for the real wage are reported in simulations). Note that real wages actually fell by about a fifth in the emerging market crises surveyed by Calvo and colleagues.

If the adverse balance sheet effects lead substantial bank closures, they may have prompt effect on the efficiency of current production as well as the level of aggregate demand. In other words, Calvo effect may also be operative. If aggregate demand falls in line with supply, employment would remain at B. But employment will increase if demand falls less than supply. This is illustrated in the figure, where the level of employment and real wage at the point labelled MZ is obtained by integrating under MPL' (not MPL) from O to C assuming that the level of demand remains the same despite the fall in productivity.

Clearly, when combined with a contraction of aggregate demand and sticky product prices, the crisis-driven fall of productive efficiency will lead to greater collapse of employment than the pure supply-side view would suggest.

V. More tools?

The dilemma for the monetary policy arises from having two objectives – to strengthen the currency and protect the economy – with only one instrument, i_1. Tinbergen's principle would suggest looking for another policy instrument.

What if tight money is complemented by an easing of fiscal policy? The logic in support of this is straightforward. If fiscal policy is used to stabilise aggregate demand in the way that exports would have (if they had time to adjust), the demand effects can be avoided. It is no surprise that IMF policy targets for fiscal *tightening* in the midst of the East Asia crisis attracted serious criticism.[20] In fact, as Fischer (2001, Chapter 1, p. 15) notes in his Robbins Lecture on 'The First Financial Crises of the Twenty First Century', "The internal debate over appropriate fiscal policy, both within the staff and with the Board [of the IMF], intensified as the crisis worsened, and as outside criticisms increased. By early 1998, budget targets began to be eased".

[20] See for example Stiglitz (1999).

More broadly, China and its neighbours now seem to believe that massive reserve accumulation at a national will insure against Sudden Stops. Is this in fact a guarantee of low sovereign spreads: or will it call for the further development of regional or global liquidity insurance schemes?

The use of capital controls, as in India (ex ante) or Malaysia (ex post), is another possible substitute for tight money in a crisis of confidence. The US stock market uses circuit-breakers to check self-fulfilling runs amplified by automatic sell orders; and London capital markets use temporary outflow controls as bulkheads to limit outflows from pension funds for example. So temporary controls are not necessarily inconsistent with developed financial markets.

VI. Conclusion

The insight of ABB was to trace the impact of devaluation on domestic investment and the supply-side of the economy via the balance sheets of credit-constrained firms. Their account of a Fisher effect for emerging markets helps to explain why emerging markets should be possessed by a "Fear of Floating". But the supply-side economic contraction they describe would surely come later and last longer than the V-shaped recession observed in SE Asia. To account for the latter, the Phoenix miracle account of CIT involves a temporary supply-side contraction of TFP driven by high interest rates[21].

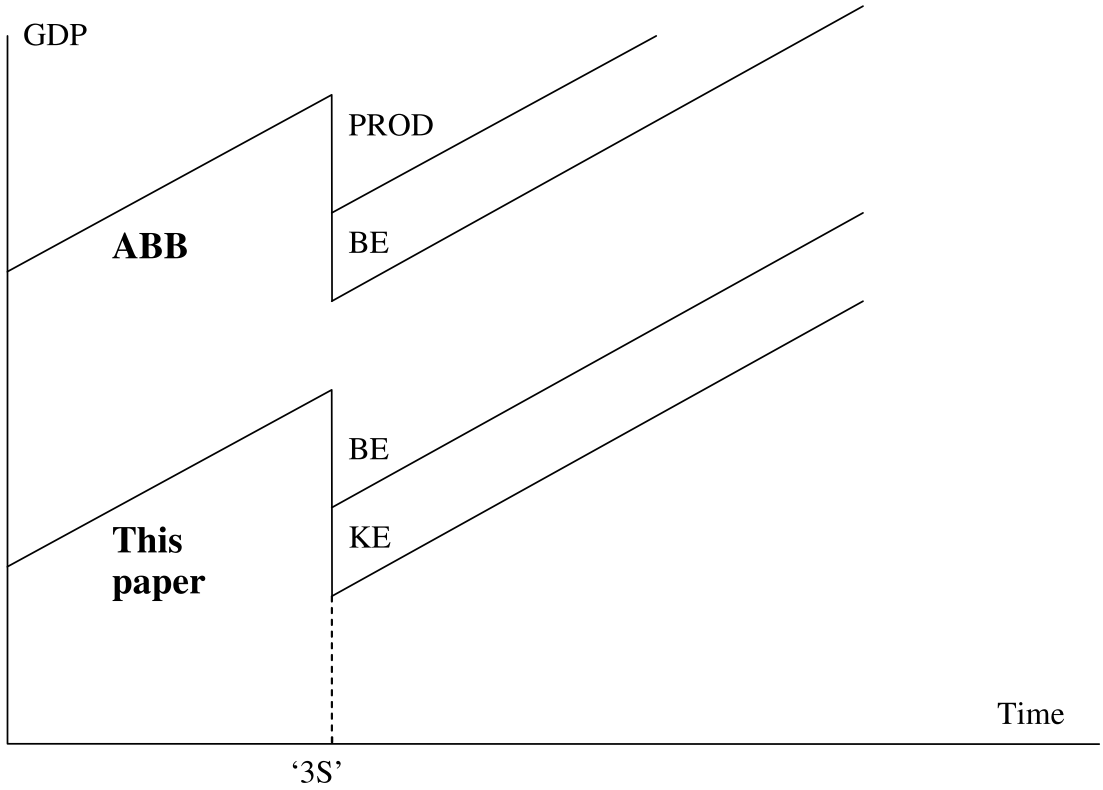
New Keynesian analyses stress the role of deficient demand for non-traded goods. National income might be adversely affected if supply-side disruptions prevent *exports* substituting for falling investment orders which could lower demand for non-tradables. But the impact of balance sheet effects in the *non-traded goods* sector[22] would have a more direct impact. .

[21] In their supply-side account of the Mexican data, Kehoe and Ruhl find it necessary to postulate an even greater role for TFP contraction: they assume a crisis-induced permanent fall in TFP of more than ten percent.

[22] (IADB, 2004 Figure 4.7, p.53 shows that in Argentina and Uruguay 70% of the liabilities of small business in the non-traded sector were dollarised).To capture this, one could put a Fisher effect into the model of Obstfeld and Rogoff (1996) outlined in the Appendix.

As for the trigger for crisis, ABB typically ascribe the Balance Sheet effects in SE Asia as the unfortunate consequence of idiosyncratic negative productivity shocks: but the coincidence of the crises in the countries concerned must throw some doubt on this interpretation. The Fisher effects may be triggered by the Sudden Stops in capital flows described in Calvo, Izquierdo and Talvi (2003).

Figure 9 Two views of crises in Emerging Markets



The resulting view of emerging market crises – and how it compares with that of ABB – is indicated in broad brush fashion in Figure 9.

The view developed in this paper is of a synchronised capital market shock affecting several open economies more-or-less simultaneously, as indicated by the symbol 3S indicating the Systemic Sudden Stop as described by Calvo et al. (2003). The resulting exchange rate collapse triggers a powerful balance sheet effect which in turn is amplified by a Keynesian multiplier. Growth resumes thereafter but at a lower level. For ABB the exchange rate collapse is attributed to adverse idiosyncratic productivity shocks which by definition reduce supply: but their impact is sharply amplified by the associated balance sheet effects.

What of the Indian experience? Does it not suggest that successful stabilisation of output in SE Asia might even have prevented *any* step-down in trend GDP. Could the long term reduction in potential be due to the lagged effects of low investment and bankruptcy in a severe demand recession?[23] On this interpretation, the contagion that spread from Thailand to Korea was not some irresistible strain of supply-side decline, but spreading exchange rate panic[24] which, properly handled, could have avoided the hysteresis effects of a recession. Treating the initiating shock not as a fall in productivity but as credit contraction induced by financial contagion might be more appropriate in such circumstances, as Aghion and Banerjee (2005, p.108) seem to acknowledge.

The framework proposed in this paper consists of an endogenous growth model of the supply side together with a demand recession triggered by balance sheet effects. (We have indicated how temporary loss of productive efficiency may be combined with demand failure.) Paul Krugman once remarked that each emerging market crisis seems to need a new economic model. But the pattern of V-shaped recession and damaged trend growth is common to both East Asia and in Latin America. Are they sufficiently similar to be analysed using a common framework -- of endogenous growth interrupted by a demand recession linked to liabilities in dollars?

References

Aghion, Philippe, Philippe Bacchetta and Abhijit Banerjee (2000), 'A simple model of monetary policy and currency crises', *European Economic Review* 44(4-6), 728–738.

Aghion, Philippe and Abhijit Banerjee (2005) *Volatility and Growth.*Clarendon Lectures in Economics. Oxford: Oxford University Press.

---

[23] There was, for example, a pronounced step-down in trend output in the US as result of the Great Depression.

[24] With respect to the crises in Indonesia and Korea, 'contagion seemed to play a dominant role. But the contagion hit economies with serious financial and corporate sector weaknesses.' Fischer (2001, chapter 1, p. 11).

Allen, Franklin and Douglas Gale (2007), *Understanding Financial Crises*, Oxford University Press.

Calvo, Guillermo, Alejandro Izquierdo and Ernesto Talvi (2006) "Phoenix Miracles in Emerging Markets: recovering without credit from systemic financial crisis" NBER WP 12101, Cambridge, MA:

Céspedes, Luis, Roberto Chang and Andrés Velasco (2003), 'IS-LM-BP in the Pampas', *IMF Staff Papers* 50(Special issue), 143–156.

Cook, David and Michael B. Devereux (2006), "Accounting for the East Asian Crisis: A Quantitative Model of Capital Outflows in Small Open Economies", Journal of Money, Credit, and Banking, Vol. 38, No. 3.

Diamond, Peter (1982), "Aggregate Demand Management in Search Equilibrium", *Journal of Political Economy*, Vol 90 (5), pp 881-94.

Diamond, D. and P. Dybvig (1983), 'Bank Runs, Deposit Insurance and Liquidity", Journal of Political Economy, 91, 401-419.

Fischer, Stanley (2001), "The International Financial System: Crises and Reform", The Robbins Lectures, LSE, London.

IADB (2004), *Unlocking credit. economic and social progress in Latin America*: 2005 Report, Washington DC.

Kehoe, Timothy J.and Kim J. Ruhl (2007) "Sudden Stop, Sectoral Reallocation and the Real Exchange Rate'. Paper presented at a conference on 'New Perspectives on Financial Globalisation' Research Department, IMF April, 2007.

Krugman, Paul (1999), 'Analytical Afterthoughts on the Asian Crisis', mimeo, MIT, available at http://web.mit.edu/krugman/www/MINICRIS.htm

Miller, Marcus, Javier Garcia Fronti and Lei Zhang (2006) "Supply shocks and Currency crises: the policy dilemma reconsidered". WERP No. 760
Moheeput, Ashwin (2007) "Monetary Arrangements, Bank Runs and Currency Crises in a Financially-Liberalised EME", mimeo, University of Warwick

Obstfeld, Maurice and Kenneth Rogoff (1996) *Foundations of International Macroeconomics,* Cambridge, MA: MIT Press

Stiglitz, Joseph (2006) *Making globalisation work*, London: Penguin, Allen Lane Norton.

Stiglitz, Joseph and Bruce Greenwald (2003) *Towards a New Paradigm in Monetary Economics*. Cambridge,UK:Cambridge University Press

Temin Peter(1976) *Did Monetary Forces Cause the Great Depression?* New York: Norton

Talvi, Ernesto (2006), 'On Sudden Stops and Output Collapses', presented at "Global Imbalances and Risk Management – Has the Center Become the Periphery?", Madrid, May

Williamson, J. (1999) 'Implications of the East Asian Crisis for Debt Management'. Paper presented to a conference on "External Debt Management", organized by the Reserve Bank of India, the Indian Ministry of Finance, and the World Bank, Kovalam, Kerala.

Woodford, Michael, (2003), *Interest and Prices.* Princeton NJ: Princeton University Press.

APPENDIX:

New Open Economy Macroeconomic model of Small Open Economy

In their New Keynesian model of a SOE, Obstfeld and Rogoff (1996, chapter 10.2) assume that the output of non-traded goods with one-period sticky prices is *demand determined* in the short run. Because they have no Fisher effect in the model, they find that unanticipated devaluation *expands* the demand for and production of nontraded goods. So this is no model of crisis! Adding a Fisher effect in the non-traded sector could presumably reduce demand and, if sufficiently strong, induce contraction after currency devaluation (much as in the traded-good-only model of ABB where export demand is constant).

The positive sign of the Keynes effect in the SOE is derived as follows.
Assume the representative consumer maximises

$$U = \left(\gamma \log C_{T,1} + (1-\gamma)\log C_{N,1} + \log\frac{M_1}{P_1} - \frac{k}{2} y_{N,1}^2 \right)$$
$$+\beta\left(\gamma \log C_{T,2} + (1-\gamma)\log C_{N,2} + \log\frac{M_2}{P_2} - \frac{k}{2} y_{N_2,1}^2 \right) + \ldots$$

where

C_T is consumption of traded good (an endowment);

C_N is consumption of non-traded good;

P is a price index, $P = P_T^{\gamma} P_C^{1-\gamma}$;

subject to period-by-period budget constraints written as

$$P_{T,1}C_{T,1} = \left(P_{T,1}\left(1+r\right)B_1 + M_0 + P_{N,1}y_{N,1} + P_{T,1}\bar{y}_T - P_{T,1}T_1 - P_{T,1}B_2 - M_1 \right)$$
$$P_{T,2}C_{T,2} = \left(P_{T,2}\left(1+r\right)B_1 + M_1 + P_{N,2}y_{N,2} + P_{T,2}\bar{y}_T - P_{T,2}T_2 - P_{T,2}B_3 - M_2 \right)$$

Since bonds are denominated in tradables and the international bond rate equals the rate of time preference (i.e. $\beta(1+r)=1$), we obtain the following first order conditions:

(1) $$\frac{\partial U}{\partial B_2} = -\frac{\gamma}{C_{T,1}} + \frac{\gamma\beta(1+r)}{C_{T,2}} = 0 \qquad \rightarrow \qquad C_{T,1} = C_{T,2}$$

[Demand for tradables]

(2) $$\frac{\partial U}{\partial M_1} = -\frac{\gamma}{P_{T,1}C_{T,1}} + \frac{\gamma\beta}{P_{T,2}C_{T,2}} + \frac{1}{M_1} = 0$$

[Money demand and arbitrage]

(3) $$\frac{\partial U}{\partial C_{N,1}} = \frac{1-\gamma}{C_{N,1}} - \frac{\gamma}{C_{T,1}}\frac{P_{N,1}}{P_{T,1}} \rightarrow P_{N,1}C_{N,1} = \frac{1-\gamma}{\gamma}P_{T,1}C_{T,1}$$

[Expenditure Pattern]

(1), (2), (3) implies money market equilibrium as follows

(4) $$M_1 = \frac{P_{T,1}C_{T,1}}{\gamma\left(1-\left(\frac{\beta P_{T,1}}{P_{T,2}}\right)\right)} = \frac{P_{N,1}C_{N,1}}{(1-\gamma)\left(1-\left(\frac{\beta P_{T,1}}{P_{T,2}}\right)\right)}$$

Log-linearising (4) as in O/R p. 693 yields the following simple specification of the evolution of money and traded-good prices – and so the exchange rate:

(5) $$m = p_T + \frac{\beta}{1-\beta}\left(p_T - \overline{p_T}\right) = c_N + \frac{\beta}{1-\beta}\left(p_T - \overline{p_T}\right).$$

(This can be obtained by setting $\varepsilon = 1$ in equation (96) of O/R)
So for an unanticipated monetary shock, $m = p_T = c_N = e$.

Interest Rate and Business Cycles in a Credit Constrained Small Open Economy*

Sarquis J. B. Sarquis†
London School of Economics / Financial Markets Group (UK)
and
Graduate School of Diplomacy / Instituto Rio Branco (Brazil)

August 12, 2008

Abstract

Empirical evidence suggests that, in contrast to developed countries, emerging economies face exogenous world real interest rates that are countercyclical, lead macroeconomic cycles and generate excessive real macroeconomic volatility. Standard small open economy models, including extensions with working capital constraints, have hardly matched these regularities, due to weak propagation mechanisms. I develop a model of a relatively impatient, credit constrained small open economy that features a permanently binding collateral constraint on net foreign liabilities. The constraint imposes an endogenous relationship between these liabilities and the capital which serves as collateral. The model reveals considerable propagation of interest rate shocks, matching impulse responses, second moments and the negative serial correlation between output and lagged interest rate. Responses to those shocks have realistic growth persistence, consistent with recession patterns. In line with the evidence, they account for over 20% of output volatility, with significant and persistent effects, beyond those of productivity shocks, over business cycles.

Key words: Business Cycles, Real Interest Rate, Credit Constraint, Collateral, Small Open Economy and World Interest Rate.
JEL Classification: F41, E32, E44, F32, G15.

*I am grateful to Francesco Caselli and Nobuhiro Kiyotaki for their guidance and support. I also thank Gianluca Benigno, Charles Goodhart, Bernardo Guimarães, Alex Michaelides, Ashley Taylor and participants of the Economics Department and the Financial Markets Group (FMG) workshops at LSE, the 2007 North America Summer Meetings of the Econometrics Society and the 2007 ESRC "Cycles, Contagion and Crises" Conference, for helpful comments and suggestions. The paper benefits as well from feedback of members of the Panel who selected an earlier version as the recipient of the LSE and GAM Gilbert de Botton Award in Finance Research 2007.

†Email address: s.j.sarquis@lse.ac.uk . Updated versions of the paper can be downloaded from http://personal.lse.ac.uk/sarquis.

1 Introduction

In the last two decades two phenomena in emerging economies have intrigued both macroeconomists and financial economists. First, these economies have registered greater real aggregate volatility than advanced economies. Second, they have faced real interest rates in international financial markets that are higher and more volatile in absolute terms than those of advanced economies.

This paper contributes to our understanding of the causal relationship between these phenomena in both empirical and theoretical dimensions. First, I provide new empirical evidence in support of the crucial role of exogenous world real interest rate shocks in emerging economies' business cycles. Second, I formulate a dynamic, general equilibrium model of a credit constrained small open economy (CCSOE) which matches most empirical regularities related to the impact of real interest rates on emerging economies. Within this model, in line with the empirical evidence, world interest rate shocks, independent of total factor productivity shocks, have systematically significant effects over business cycle frequencies, explaining over 20% of output volatility.

As argued by Calvo et al. (1996), international financial factors can be a key driver of real interest rates of emerging economies over business cycle frequencies. They can determine overall conditions of world credit markets, in particular those of emerging markets, beyond the impact of domestic variables. Recent empirical studies indicate that, although subject to endogenous forces (via the country spread), real interest rates are essentially driven by exogenous global credit shocks. These shocks propagate domestically through the interest rate by the means of credit frictions. Using a richer representation of global credit markets, Sarquis (2006) shows that such shocks can, in the case of Brazil, account for even larger shares of macroeconomic variability. The latter increase from just over 20% (as found in Uribe and Yue, 2006) to over 50%, if the representation of global markets includes, beyond a benchmark US interest rate, other exogenous variables that are able to reflect changes in perceptions towards risks and uncertainties at the core of those markets.

In this paper I use a VAR analysis of the Brazilian economy to identify the main empirical regularities that a model of emerging economies should be able to reproduce. Adding to previous work which clarifies that real interest rates are exogenous, counter-cyclical and lead cycles (e.g. Agénor and Prasad, 2000, Neumeyer and Perri, 2006, and Uribe and Yue, 2006), I show that international real interest rates have the ability to generate recessions and recoveries that are consistent with the dynamics of growth persistence. In particular, I highlight that the responses of output and consumption to world real interest rate shocks are hump-shaped. Overall, I argue that any theoretical model of such features requires strong propagation of these shocks.

The credit constrained small open economy (CCSOE) model that I propose contains two main innovations compared to standard small open economy business cycle models. First, the stock of net foreign liabilities is endogenously constrained by the accumulating capital stock, which works as collateral in the sense of Kiyotaki and Moore (1997). Second, the economy's representative agent is relatively impatient. Thus, the economy finds itself not only in a negative net foreign asset position in steady state, but also facing a permanently binding credit constraint. These two complementary assumptions are key to the propagation mechanism.

Furthermore, I set additional assumptions with regard to consumption preferences and adjustment costs to labour and investment only to control at the margin the corresponding responses. These assumptions are irrelevant with regard to the major qualitative implications of the model and do not change the bulk of the quantitative results of its simulations. Following Jaimovich and Rebelo (2006), preferences are set in more general terms, combining characteristics of the two most widely used formulations in models of real business cycles and of small open economies. They help to obtain a more realistic timing for the troughs in the hump-shaped responses. Standard adjustment costs to labour and investment are introduced merely to control the excessive variability of these variables, especially of the latter.

By an impatience hypothesis and by the implication of a permanently binding collateral constraint, the model captures two aspects of the financial integration of emerging economies with the world economy. First, it stresses their structural weakness in promoting domestic savings and/or financial deepening - as observed in Latin American countries over the last decades - and consequently their dependence on foreign credit. Second, it focuses on frictions in international financial markets which, by the means of collateral, provide foreign creditors and investors with some discipline or rationing over the country's international financial exposure.

The key propagation mechanisms are associated with growth persistence in output and consumption. Such persistence results from the dual role of accumulating capital, as both collateral and production factor. The reliance on foreign finance, by the means of collateral, gives the economy an additional benefit in avoiding dramatic falls in the capital stock, in response to adverse shocks. Therefore, when investment is curtailed, it recovers monotonically afterwards, while consumption adjusts smoothly in hump-shaped form. The propagation mechanisms that underlie capital and consumption growth dynamics are interrelated, coinciding with a variable premium between the marginal product of capital and the interest rate. Recessions and recoveries are triggered by premium levels respectively below and above the steady state value. In response to adverse interest rate shocks, the premium declines and the marginal product of capital becomes comparatively low, creating disincentives to invest and consume. Recessions are prolonged, as the interest rate still declines above its steady state value. Meanwhile, the marginal product of capital and the corresponding premium rise, prompting recoveries.

Calibrated to Brazil, the model's simulations match well the empirical evidence, structure of second moments and the negative correlations between output and lagged interest rates. They also replicate the empirical responses of output and consumption to interest rate shocks, displaying hump-shapes with troughs that occur realistically around 3 quarters after the shock. Overall, interest rates have significant effects in determining macroeconomic volatility. Over business cycle horizons these effects are not dampened by productivity shocks with realistic or similar standard deviations. Interest rate shocks increment the overall variability of real aggregates. Moreover, due to a persistent impact on consumption growth dynamics, they generate a realistically higher volatility of consumption relative to output.[1]

In all these accounts, thanks to its superior propagation properties, the CC-SOE model seems to outperform other small open economy models in addressing systematically key facts related to emerging economies' business cycles. Motivated by the *real* business cycle approach inaugurated by Mendoza (1991) for SOEs, most models do not impose a financial friction in the endogenous sense here proposed. They are standardly characterized by the use of one of the "ad hoc" hypotheses discussed in Schmitt-Grohé and Uribe (2003) - to close the SOE model and to render it stationary - that are essentially equivalent, sharing eventually the same poor propagation mechanism in consumption dynamics.[2]

Although widely recognized empirically as a potentially important mechanism for transmitting international shocks, the world real interest rate can only play a limited role in standard SOE models. That is why alterations to the standard framework, including to shock specifications, have been proposed. Blankenau, Kose and Yi (2001) reverse the standard RBC methodology and back out from the model a specification of shocks in which about one-third of Canada's output volatility could be explained by interest rate shocks. But the volatility required for these shocks is about 8 times the volatility of total factor productivity shocks. Neumeyer and Perri (2006) and Uribe and Yue (2006) incorporate, among other features, a working capital constraint. The two models perform relatively well in matching certain regularities, respectively the negative correlation between output and lagged interest rates and the responses to interest rate shocks. However, they depend on additional assumptions to generate the appropriate propagation in each case. Merely incorporating a working capital constraint would not suffice to attain the results. The former explicitly acknowledge the reliance on country-spread shocks that are independent from

[1] See for instance Agénor and Prasard (2000) and Kydland and Saragoza (1997).

[2] Schmitt-Grohé and Uribe (2003) address a variety of stationary assumptions in infinite-horizon models (endogenous discount factor, debt-elastic interest rate premium and portfolio adjustment costs). Alternatively, Blanchard (1985) and Yaari (1965) set stationarity assumptions in an overlapping generations model.

world interest rate shocks, but induced by negative productivity shocks. The latter use a VAR estimated equation of the country-spread to close the model. Overall, to reconcile theory and evidence, most SOE models have to rely on unrealistically high volatility of interest rates and/or to assume a given negative correlation between the latter and productivity (Oviedo, 2005).

While intrinsically enhancing the propagation of interest rates, the CCSOE model also overcomes anomalies of standard SOE models: a low and insignificant correlation between output and the net export ratio to GDP and a lack of serial auto-correlation in investment. It recovers dynamic properties of the closed-economy real business cycle models that, despite being empirically imperative, have been widely neglected within the standard SOE framework. The latter is nested by the CCSOE as the extreme case of a credit frictionless economy.

The CCSOE model inherits some features of Kiyotaki and Moore (1997) or Kiyotaki (1998), but differs from their's in the following important aspects. First, as a SOE model, it incorporates exogenous interest rate shocks. Second, by the same token, the economy has a larger variety of routes for adjusting to shocks, namely via reversals in the current account. Third, capital rather than land acts as collateral. Since the former asset evolves according to an aggregate accumulating dynamics, the economy is widely subject to the adjustment processes associated with changes in net liabilities. Fourth, combining access to foreign finance and aggregate collateral formation enhances the propagation mechanism.

Arellano and Mendoza (2002), Chari et al (2005) and Kocherlakota (2000) also use foreign credit constraints in SOE models.[3] They however stop short of solving a dynamic stochastic general equilibrium (DSGE) model in which both capital and foreign liabilities are endogenously accumulated. Moreover, they do not attempt to match real data, namely business cycle statistics and responses. Mendoza (2006) is the closest model, but has different motivation and hypotheses. First, I am concerned with more regular business fluctuations rather than occasional sudden stops. Second, closing the SOE model by a permanently binding constraint, I rule out the nonlinearity associated with a slack credit constraint and do not need to add another assumption only to render the system stationary.

Responses to real interest rate shocks in emerging economies share similar features to the responses to domestic monetary policy shocks as frequently documented for advanced economies. In both cases their humped-shape indicate the need to bring growth persistence to the core of the propagation mechanism of macroeconomic models. A modelling strategy based on nominal rigidities, as in Christiano et al. (2005), could be extended to SOEs and complement the explanation offered in this paper. Abstracting from nominal rigidities and domestic

[3] Caballero and Krishnamurthy (2001) also develop a capital collateral on foreign borrowing in a three-period model.

monetary policies allows us however to properly address the *real* propagation mechanisms - with credit frictions - of world *real* interest rates in emerging economies.

Section 2 presents the empirical evidence and regularities. Section 3 describes the model. Section 4 deals with the calibration. Section 5 and 6 contain the main results, while comparing model simulations with the empirical evidence. Section 7 discusses the intuition behind the propagation mechanism of the model. Section 8 explores the robustness of the model, and section 9 concludes.

2 The Evidence on Real Interest Rate Shock

In order to revisit and to clarify the evidence on the role of real interest rate shocks in emerging economies, as well as to set an empirical benchmark for the model of Section 3, I conduct a VAR exercise on Brazil for the period 1994-q2 to 2005-q4. Brazil is one of the major developing economies, being responsible for a significant share in (the global segment of) the so-called emerging markets. The period of analysis covers most of the Brazilian experience of closer integration with global financial markets, common to similar emerging economies.[4] Moreover, the essence of the Brazilian evidence coincides with the cross-country evidence that has been put forward for emerging economies.[5]

The VAR representation is one of a simple SOE economy, in which the following endogenous variables are included: output, hours, consumption, investment and net trade to GDP ratio.[6] They enter, with the same order, the vector y. Real interest rate r, is also included. It is obtained as the difference between the nominal rate of US 3-month Treasury bonds and the corresponding expected US inflation.[7] The variables are in logs, except for the trade ratio and interest rate, and enter the VAR in levels. Additional details on the data are described in the Appendix.

The VAR is set with only one (1) lag. This choice is guided by information criteria tests (Akaike, Final Prediction Error, Hannan-Quinn and Schwartz), allowing for a maximum of 12 lags. Apart from time trends, which are added in the equations of output, consumption and investment, the unrestricted VAR representation is the following:

$$\begin{bmatrix} y_t \\ r_t \end{bmatrix} = \begin{bmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{bmatrix} \begin{bmatrix} y_{t-1} \\ r_{t-1} \end{bmatrix} + u_t.$$

2.1 Exogeneity and significance

VAR Granger causality tests and variance decomposition[8] analysis support firmly the hypothesis that interest rate can be treated as an exogenous process,

[4] See Kose, Prasad, Rogoff and Wei (2006) for a reappraisal of financial globalization and its relationship with emerging economies.

[5] See e.g. Neumeyer and Perri (2006) and Uribe and Yue (2006).

[6] This representation is similar to the one used by Uribe and Yue (2006), but includes also hours and consumption. The latter and Neumeyer and Perri (2005) treat Brazil as a SOE, whereas Kanczuk (2004) constructs a closed-economy model to study real interest rates and the country's business cycles.

[7] Expected inflation is given by the estimate of an autorregressive process with 8 lags.

[8] Huh (2005) makes an interesting point on the possible limitations of Granger causality tests especially for VAR analyses of SOEs, suggesting the use of variance decompositions to further assess the hypothesis of exogenous (foreign) variables.

independent from the representation of the macroeconomy given by the five endogenous variables in y. Granger causality tests, which result from the unrestricted VAR, are shown in Table 1.

Table 1. Granger causality tests

H_0 :	p-value
$y \rightarrow r$	0.4660
$r \rightarrow y$	0.0023

Variance decompositions are calculated for four specifications: unrestricted VAR; restricted VARs with r ordered last and first; and the VAR with exogenous r. The latter obtains by setting $c_{21} = 0$, and therefore the effects of interest rate shocks do not depend at all on the ordering.

Table 2 reports the variance decompositions under the above specifications, at 4, 8 and 20 quarters after the shock. The unrestricted VAR decompositions are not shown, as they coincide with those of the restricted VAR with interest rate ordered last. Overall, all specifications indicate a strong degree of exogeneity for the interest rate and a significant effect of interest rate innovations on economic activity over business cycle frequencies. In line with Granger tests, there is no empirical support for reverse causality. Interest rates appear to explain around 30% of the variability of output and investment, a higher proportion (40%) of consumption's, and about 20% of trade movements.

Table 2. Variance decompositions due to interest rate in VAR models

Variable	Per cent (%) in VAR models								
	restricted with endogenous interest rate						with exogenous interest rate		
	$[y_t\ r_t]'$			$[r_t\ y_t]'$			$c_{21} = 0$		
	quarters			quarters			quarters		
	4	8	20	4	8	20	4	8	20
Interest rate	90	88	87	99	97	96	100	100	100
Output	23	26	25	22	25	25	26	37	37
Hour	4	3	3	3	2	3	4	4	3
Consumption	34	41	41	32	41	41	35	51	53
Investment	26	32	32	20	28	28	23	38	40
Trade	8	18	17	9	19	18	6	19	22

Note: Unrestricted VAR obtains the same decompositions of the restricted VAR with interest rate ordered last.

These results are consistent with those in the empirical literature that support the hypothesis that - to a large extent - exogenous shocks to emerging economies' interest rates are key to the understanding of their business cycles. In a panel VAR including seven emerging countries (Argentina, Brazil, Ecuador, Mexico, Peru, Philippines and South Africa), Uribe and Yue (2006) find that on average about 20% of movements in aggregate activity is explained by disturbances in the US real interest rate. This result has been reinforced by Canova (2005), demonstrating the relative importance of US monetary shocks. Sarquis (2006) finds, within a more comprehensive representation of world credit markets, that by including innovations to other US financial variables, such as changes in term spreads in US interest rates and in premia for Moody's BAA corporate bonds (over Moody's AAA corporate bonds), over 60% of the variability of Brazil's real interest rate and over 50% of the country's output volatility could be accounted by exogenous financial factors that are at the core of global credit markets. He shows that these factors, affecting uncertainties and risk perceptions, transmit into the real economy through via the interest rate and its corresponding financial acceleration mechanism.

2.2 Propagation, persistence and volatility

The solid lines in Figure 1 are the estimated responses of *all* endogenous real variables resulting from the VAR with exogenous interest rate shocks. Dotted lines represent 95% confidence intervals. The other VAR specifications have identical responses and confidence intervals.

The estimated responses feature the dynamics of growth persistence. Output, consumption and investment responses conform with recessions (recoveries) in which a drop (rise) in one of these real aggregates is followed subsequently by another.[9] The troughs occur after two to four quarters, and the variables appear to return to their pre-shock levels after six to twelve quarters. The responses indicate that interest rates are effectively countercyclical and leading the cycle.[10] Furthermore, the trade ratio response denotes a positive adjustment in the current account, which can be associated to larger surpluses or smaller deficits. It also seems to be sluggish, with considerable persistence and picks around the 5th quarter. All the responses are thus marked by hump-shaped patterns. They summarize the propagation of interest rate shocks, with pronounced and magnified recessions. Their sluggishness and growth persistence dynamics is analogous to those found in studies of the impact of productivity shocks and monetary shocks in closed-economy business cycles.[11]

[9] Note that, given the one-sided nature of the responses, the likelihood of misrepresented recessions is only 2.5%.

[10] The counter-cyclical feature of real interest rates might not be exclusive to emerging economies. For instance, King and Rebelo (1999) and Stock and Watson (1999) report counter-cyclical evidence for the US, which was provoked perhaps by different causes, such as a less stable monetary policy before the nineties.

[11] See, for instance, Cogley and Nason (1995) and Christiano et al. (2005), respectively.

Due to such a strong propagation, correlations between output and lagged interest rate are persistently negative, peaking in absolute terms from two to three quarters after the shock. Otherwise, in the absence of such a propagation, the effects of interest rate shocks would be weakened and dampened by dominant shocks over business cycle frequencies, as in most SOE models. In the CCSOE model sustained and significant effects of interest rate shocks - among other possible sources of macroeconomic fluctuations - obtain mainly because of the strength the propagation mechanism derived from its credit frictions.

Interest rates emerge thus as a potentially considerable source of the excessive macroeconomic volatility of emerging economies. Over the period of analysis, Brazil reveals a standard deviation of the HP-filtered GDP series (in logs) that is 1.78 higher than the US counterpart. At the same time the volatility of Brazil's real interest rate is about twice that of the US rate.

Moreover, interest rates can be an important force behind the high volatility of consumption, relative to output, among emerging economies. As already revealed in Figure 1, interest rate shocks are prone to generate an excess in the deviation from steady state in consumption response vis-à-vis output response. This fact is further corroborated by the unconditional second moments of the Brazilian real data, shown in Table 5 with simulated moments.[12]

2.3 Theoretical challenges

Following the VAR analysis of the Brazilian case, the overall evidence on interest rate and business cycle in emerging economies suggests that:

(a) shocks to the economy's real interest rate can be mainly - or to a good extent - exogenous and certainly sustained and significant over business cycle frequencies;

(b) output, consumption and investment respond to these shocks in hump-shaped form, in agreement with the persistent feature of recessions and recoveries;

(c) the real interest rate is unambiguously counter-cyclical and leads the cycle, with strong propagating forces, regardless of the effects of other sources of fluctuations;

(d) correlations between output and lagged real interest rates are increasingly negative during recessions, picking up in absolute terms around the trough of output responses;

(e) consumption reacts relatively more than output to interest rate shocks;

(f) current account responses are also sluggish, in a hump-shaped pattern, and at least as counter-cyclical as in advanced economies.

[12] Table 5 shows an excess of 29% in consumption's standard deviation relative to output's. Neumeyer and Perri (2005) calculate a similar excess of 24% for a shorter period.

Most of these challenges are not necessarily found in advanced economies and cannot be explained in an integrated way by available SOE models. To address them all, it appears that a model must have ideally four features: a stronger propagation of interest rate shocks, particularly with respect to consumption's intertemporal dynamics; a propagation mechanism by which recessions are aggravated by negative growth persistence, and by which responses to interest rate shocks conform to hump-shapes; as a result of the above two features, a good matching of standard second moments of data statistics, showing in particular the ability of interest rate to generate excessive macroeconomic variability and more realistic comovements, autocorrelations and relative deviations of the series; and, finally, a good replication of (the dynamic pattern of) correlations between output and interest rate at lags that are empirically meaningful.

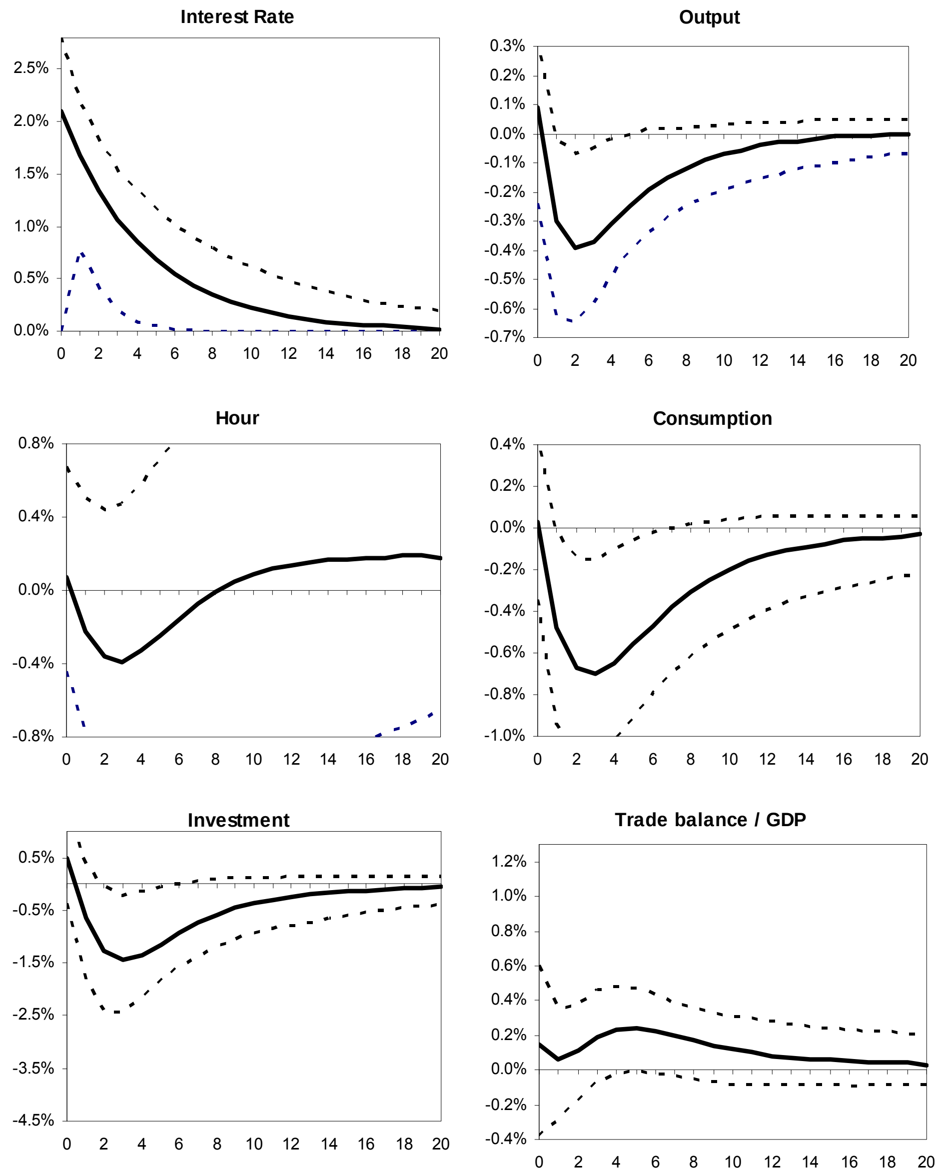


Figure 1. - VAR impulse responses. Solid lines are VAR responses to interest rate shocks. Dotted lines represent 95% confidence intervals. The vertical axis shows deviations from steady state. Units on the horizontal axis are quarters.

3 The Model Economy

The model economy has a single homogeneous good and is populated by a single representative agent. It faces an exogenous world economy against which it has net foreign liabilities (sustained recourse to foreign financing) paying an exogenously determined and variable gross rate R_t. The lower bound of this gross interest rate is given by a benchmark international rate, R_t^*, so that $R_t \equiv 1 + r_t \geq 1 + r_t^* \equiv R_t^*$, and $R \equiv E_t R_t \geq E_t R_t^* \equiv R^*$.

I assume the representative agent is relatively less patient than the world economy's counterpart, whose discount factor $\beta^* \equiv R^{*-1}$, as usually set. Therefore, the CCSOE's discount factor is lower than the rest of the world's, that is $\beta < \beta^*$.[13] Such a (parametric) relative impatience in relation to the rest of the world is similar to Paasche's (2001). It is also analogous to the impatience gap between heterogeneous agents assumed in the closed-economy models of Kiyotaki and Moore (1997) and of Carlstrom and Fuerst (1997). Moreover, it is in line with the heterogenous cross-country empirical evidence, which indicates a positive correlation between β and wealth, as in Becker and Mulligan (1997).

The relative impatience gives the model steady state properties that preclude the use of one of the assumptions that are set standardly only to close SOE models and to render them stationary (Schmitt-Grohé and Uribe, 2003). As long as $R\beta < 1$, a steady state equilibrium obtains in which the collateral constraint permanently binds.[14] By setting the impatience hypothesis, I explicitly assume a prior about the emerging economy I want to model: it would like to borrow as much as it could, but its ability to borrow is constrained by its own performance and the interest rates it pays in global markets.

The representative agent maximizes her life-time utility defined as in Jaimovich and Rebelo (2006),

$$U = E_t \sum_{s=t}^{\infty} \beta^{s-t} u(C_t, L_{t,}, J_t), \tag{1}$$

where

$$u(C_t, L_{t,}, J_t) = \frac{(C_t - aH_t^{\theta} J_t)^{1-\eta} - 1}{1-\eta} \tag{1a}$$

$$J_t = C_t^b J_{t-1}^{1-b} \tag{1b}$$

[13] In line with the SOE assumption I do not model the rest of the world, which would correspond to the case of an unconstrained SOE representative agent model. Even if it is assumed a patient economy, as it lends to the CCSOE, the ratio of its net foreign assets to capital would be close to zero. Accordingly, $\beta^* R^* = 1$.

[14] Note that the impatience hypothesis ($\beta < \beta^*$) is implied by $\beta R < 1$ and $R > R^*$. Arguably, to set $\beta < \beta^*$ appears to be as "ad hoc" as to set $\beta = \beta^*$ or even $\beta > \beta^*$.

and

$$L_s + H_s = 1 \tag{1c}$$

L and H stand for leisure and hours worked. While C_t denotes current consumption, it can be shown that J_t refers to an index that tracks the consumption path. It could be interpreted as the "underlying" consumption level, which controls the marginal substitution of leisure relative to actual consumption. The above representation of preferences nests, on the one hand, preferences of standard RBC models, such as King, Ploser and Rebelo (1998), and, on the other hand, preferences widely used in standard SOE models, originally found in Greenwood, Hercowitz and Huffman (1988)[15]. It assures therefore a realistic approach to labour dynamics, avoiding excessively fast (short term) and excessively sluggish (long term) labour adjustments over business cycles. It can provide realistically the supply of labour with moderate wealth effects.

The agent accumulates capital and faces not only resource and technology constraints, but also a credit collateral constraint on net foreign liabilities.

The resource constraint is the following:

$$C_t \leq Z_t f(K_{t-1,}L_t) - I_t + B_t - R_t B_{t-1} - H_t \Omega(H_t/H_{t-1}) \tag{2}$$

The agent can therefore finance consumption and investment expenditures with resource to net foreign liabilities.

The function $\Omega(\ )$ defines labour adjustment costs, and I assume $\Omega(1) = \Omega'(1) = 0$ and $\Omega''(1) = \pi^h$.These assumptions are sufficient to determine the costs incurred for changes in labour, while no costs are incurred in steady state. Z_t is current total productivity. The production function is Cobb-Douglas and therefore:

$$f(K_{t-1,}L_t) = K_{t-1}^{\alpha}(1 - L_t)^{1-\alpha} \tag{3}$$

The capital accumulation is given by:

$$I_t[1 - G(I_t/K_t)] = K_t - (1 - \delta)K_{t-1} \tag{4}$$

The function $G(\ )$ represents adjustment costs to investment. Analogous to labour adjustment costs, I simply assume $G(1) = G'(1) = 0$ and $G''(1) = \pi$. The latter parameter does not affect the steady state properties of the model, but its dynamic properties. The rate of capital depreciation is given by $\delta \in [0, 1]$.

The credit collateral constraint, which always binds due to the relative impatient assumption, is:

[15] These preferences obtain, respectively, for $b = 1$ (KPR preferences) and $b = 0$ (GHH preferences).

$$B_t \leq \gamma_t[K_{t-1}(1-\delta)+\sigma I_t] \tag{5}$$

where $\gamma_s \in [0,1]$ designates the proportion of capital that is actually accounted as collateral formation. Current investment might play a role in collateral formation, by a proportion given by σ. For $\sigma = 0$ or $\sigma = 1$, the constraint would be respectively $B_t \leq \gamma_t K_{t-1}(1-\delta)$ or $B_t \leq \gamma_t K_t$. Correspondingly, either current investment would have no value in collateral formation or it would have exactly the same value as of physical capital. In the baseline calibration I set $\sigma = 0$, since it assures less volatility of investment. However, an intermediate case, in which current investment has some value should not be dismissed.

The model is subject to three disturbances that can affect the exogenous processes of productivity, interest rate and collateral formation. These processes follow a vector auto-regressive form:

$$w_t = Pw_{t-1} + \varepsilon_t \ ,$$

where $w_t = \begin{bmatrix} z_t & r_t & \gamma_t \end{bmatrix}'$and $\varepsilon_t = \begin{bmatrix} \varepsilon_t^z & \varepsilon_t^r & \varepsilon_t^\gamma \end{bmatrix}'$. Note that $z_t = \ln Z_t$ and $r_t = \ln R_t$.

The benchmark model has essentially a diagonal P, in which only independent, temporary and persistent shocks to productivity and interest rate. To explore the potential of the model's propagation mechanism of interest rate shocks, alternative (non-diagonal) specifications of P are considered in which the exogenous processes cease to be independent, while shocks remain so.

3.1 First order conditions

The problem involves maximizing the following Lagrange expression, with λ_t , $\lambda_t q_t$, $\lambda_t \varphi_t$ Lagrange multipliers:

$$\mathcal{L} = E_t \sum_{s=t}^{\infty} \beta^t \left\{ \begin{array}{c} u(C_t, L_{t,}, J_t) + \psi_t[J_t - C_t^b J_{t-1}^{1-b}] \\ \lambda_t[Z_t f(K_{t-1,} L_t) - I_t + B_t - R_t B_{t-1} - C_t - H_t \Omega(H_t/H_{t-1})] \\ +\lambda_t q_t[I_t - I_t G(I_t/K_{t-1}) - K_t + (1-\delta)K_{t-1}] \\ +\lambda_t \varphi_t[\gamma_t K_{t-1}(1-\delta) + \gamma_t \sigma I_t - B_t] \end{array} \right\}$$

Six first order conditions obtain from the Lagrangian maximization problem:

$$u_L(C_t, L_{t,}, J_t) = -\lambda_t \left[f_L(K_{t-1,} L_t) + \Omega(H_t/H_{t-1}) + \frac{H_t}{H_{t-1}} \Omega'(H_t/H_{t-1}) \right] \tag{6}$$

$$+\beta \lambda_{t+1} \left(\tfrac{H_{t+1}}{H_t} \right)^2 \Omega'(H_{t+1}/H_t)$$

$$\lambda_t = u_C(C_t, L_{t,}, J_t) - b\psi_t \left(\frac{J_{t-1}}{C_t}\right)^{1-b} \quad (7)$$

$$\psi_t = -u_J(C_t, L_{t,}, J_t) + \beta(1-b)\psi_{t+1}\left(\frac{J_t}{C_{t+1}}\right)^{-b} \quad (8)$$

$$\lambda_t = \lambda_t q_t \left[1 - G(I_t/K_{t-1}) - \frac{I_t}{K_t}G'(I_t/K_{t-1})\right] + \gamma\sigma\lambda_t\varphi_t \quad (9)$$

$$E_t\beta\frac{\lambda_{t+1}}{\lambda_t}R_{t+1} = 1 - \varphi_t \quad (10)$$

$$E_t\beta\frac{\lambda_{t+1}}{\lambda_t}\left\{\begin{array}{c} Z_{t+1}f_K(K_t, L_{t+1}) + q_{t+1}\left[1 - \delta + \left(\frac{I_{t+1}}{K_t}\right)^2 G'(\frac{I_{t+1}}{K_t})\right] \\ +(1-\delta)\gamma_{t+1}\varphi_{t+1} \end{array}\right\} = q_t \quad (11)$$

Equations (6) and (7) govern the standard intra-temporal consumption-labour substitution, except that they contain terms related respectively to labour adjustment costs and, more importantly, to (inter-temporal) deviations in the consumption path. The latter enriches the standard substitution problem in ways that are determined by (8), which sets the dynamics of the disturbances to consumption, as in Jamovich and Rebelo (2006).

The first order condition with respect to investment, equation (9) controls the movements in the shadow price of investment. This price would be constant ($q = 1$), had I assumed no investment adjustment cost and no role for current investment in the credit collateral formation.

Equations (10) and (11) are the fundamental Euler conditions. They differ sharply from the counterparts standardly found in RBC or SOE models. They both contain the relative multiplier of the credit collateral constraint, φ_t, which gives the shadow price of collateral relative to consumption. The binding of the constraint imposes a positive value for the shadow price. At the steady state, $\varphi = 1 - \beta R$. Equation (10) has been stressed in previous work, such as Chary et al. (2005), Arellano and Mendoza (2002) and Mendoza (2006). However, in their models, except to a certain degree in the latter's, the multiplier does not figure explicitly as in Equation (11). The two equations are key to the propagation mechanism that characterizes the CCSOE framework.

The model is fully described by Equations (2)-(5) and by the FOCs expressed in equations (6)-(11), as well as by the specification of the underlying exogenous processes, which I address in the next Section. I solve the model by the method of logarithmic linearization, as described in Uhlig (1999).[16]

[16] The Matlab code containing the loglinearized equations can be provided by the author. The simulations use Uhlig's toolkit of Matlab codes for analyzing nonlinear dynamic stochastic models.

4 Calibration

4.1 Parameters

The calibration is guided by Brazilian data and by restrictions imposed by the structure of the model. The baseline parameters are shown in Table 3. The average international real interest rate was around 8% per annum (therefore $r = 0.02$ on a quarterly basis) in Brazil over 1994 to 2005. According to the country's national accounts, the shares of investment and net export in GDP (I/Y and X/Y) were around 20% and 3%.

In business cycle studies on Brazil, the choice of the capital ratio (K/Y) varies between 4 (Ellery, Gomes and Sachida, 2002) and 11.6 (Kanczuk, 2004), on a quarterly basis (1 and 2.9 on an annual basis). Empirical estimates of capital and therefore of the ratio have shown similar discrepancies. I set an intermediate value, at 7.45 on a quarterly basis (1.86 on an annual basis). It is just below typical values used for the US or for sall open advanced economies, such as Canada. For the latter, since Mendoza (1991) most studies has set the ratio at 8.8 (2.2 on an annual basis).

To specify technology parameters, calculations based on income shares usually give developing countries a higher capital share (α) than found for developed economies. I use $\alpha = 0.38$, although Brazilian data would suggest a value close to 0.50. The under-estimation of labour income results likely from the use of informal and/or self-employed labour.[17] Capital and labour adjustment costs' parameters (π and κ) follow typical ranges found in the literature.

The subjective discount rate, that is implied by the discount factor, must be higher than the real interest rate in the CCSOE model. Correspondingly, the steady state premium, φ, given by $\varphi = 1 - \beta R$, is set on a quarterly basis at 0.0051. This is consistent with the hypothesis of relative impatience and with estimates of Brazil's discount factor, which usually are close to 0.9 annually - well below the estimates for the US.[18] Apart from the discount factor, the calibration of the preference parameters is similar to Jaimovich and Rebelo (2006). In particular in the baseline calibration, I use a small value for b, rendering preferences closer in spirit to the GHH preferences. This actually permits that hours react negatively in the short run to adverse interest rate shocks, as indicated in the corresponding VAR response in Figure 1.

The collateral parameter γ is set in accordance with the observed stationary of net foreign liability ratio, B/Y, as well as of the capital ratio, K/Y. It can be shown from Equations (4) and (5) that in steady state:

[17] See Golin (2002) for an overall discussion on income shares, and Caselli and Feyrer (2006) for cross-country comparisons in capital share. Shares in Latin American economies range from 0.4 to 0.5, while in advanced economies they lie between 0.2 and 0.4.

[18] Estimates by by Val and Ferreira (2001) and Issler and Piqueira (2001), under different utility specifications, support this value. Erely et al. (2002) use a similar value (0.89 annually).

$$\frac{B}{Y} \equiv \frac{K}{Y}\frac{B}{K}\gamma\left[1-(1-\sigma)\delta\right]$$

Setting $\sigma = 0$ and $\delta = 0.027$, we can recover a consistent value for γ. I roughly calculate the country's net foreign liabilities (liabilities minus assets), by adding the net foreign debt (0.28) and remaining net foreign liabilities in equity, minus the international reserves. I arrive at a net foreign liability ratio of 0.42 annually. Thus, γ is around 0.22. A lower γ would obtain had I used a narrow concept of the country's international investment position, such as the stock of merely international debt contracts.

Table 3. Baseline parameter values

Name of parameters	Symbol	Value
Rates		
international real interest rate	r	1.02
subjective rate	β^{-1}	1.0251
Preference		
discount factor	β	0.9755
elasticity of labour supply $= \frac{1}{1-\theta}$	θ	1.01
utility curvature	η	1
utility parameters	a	2.39
	b	0.15
Technology		
capital share	α	0.38
depreciation rate	δ	0.027
capital adjustment cost	π	2.6
labour adjustment cost	κ	2.0
Collateral formation		
collateral share	γ	0.2238
current investment weight	σ	0

4.2 Specification of the exogenous processes

The CCSOE model of Section 3 is subject to three exogenous processes: productivity, real interest rate and collateral formation. They are represented in vector form by

$$w_t = Pw_{t-1} + \varepsilon_t \ ,$$

where $w_t = \left[\begin{array}{ccc} z_t & r_t & \gamma_t \end{array} \right]'$.

The first two processes are conventional in SOE models. Collateral formation is here proposed as an exogenous process that represent innovations in the efficiency by which the country can provide international creditors and investors with collateral through capital accumulation. It relates to a country's ability to borrow and creditors's willingness to lend. It can result from short or long term factors at home - e.g. underdevelopment of the country's financial markets and associated weakness - or in international financial markets - e.g. segmentation of international financial markets with regard to a country or to a class of countries to which it belongs.

I examine two kinds of specifications to control the dynamics of the exogenous processes: purely independent and interest rate induced processes. In the former each process is completely, driven by its own independent shocks. In the latter kind of specifications, not only each process can be driven by its own shocks, but also real interest shocks can affect the collateral and productivity processes. I assume however that all shocks are always treated as independent, so that no correlation is assumed between them.

Overall, consistent with the empirical evidence, I assume that innovations to the interest rate are a key source of fluctuations. That is why it is imperative to analyze the specification with purely independent shocks. Furthermore, since the interest rate does not appear to suffer from a reverse causality from domestic factors, but rather to cause virtually all observable macroeconomic aggregates, it should be seen as a good candidate to drive the economy's international macroeconomic and financial conditions, beyond its impact domestically. It might trigger other international macroeconomic and financial factors that further propagate shocks beyond the (independent) interest rate self-propelled propagation mechanism.

Therefore, all the proposed specifications for P feature the restriction that $P_{2j} = 0$, for all $j = 1$ and 3. The *purely independent* specification has a diagonal matrix P, in the following general form:

$$P = \begin{bmatrix} \rho^z & 0 & 0 \\ 0 & \rho^r & 0 \\ 0 & 0 & \rho^\gamma \end{bmatrix}$$

Interest rate induced processes obtain alternatively if one or more nondiagonal elements in P that reflect an interest rate causation are nonzero; that is: $P_{i2} \neq 0$, for $i = 1$ and 3. In general, P would conform thus to:

$$P = \begin{bmatrix} \rho^z & \rho^{zr} & 0 \\ 0 & \rho^r & 0 \\ 0 & \rho^{\gamma r} & \rho^\gamma \end{bmatrix}$$

The values I attribute to P in the model's simulations are either derived directly from the coefficient estimates of the VAR or indirectly motivated by them. Table 4 shows parameter estimates of the unrestricted VAR of Section 2. Alternative VAR specifications give almost identical estimates.

Table 4. Parameter estimates of the unrestricted VAR

	Dependent variable in VAR					
	$\hat{y}_t$	$\hat{h}_t$	$\hat{c}_t$	$\hat{i}_t$	$\hat{x}_t$	$\hat{r}_t$
$\hat{y}_{t-1}$	0.271	0.061	-0.163	0.479	0.522	-0.127
	(0.154)	(0.230)	(0.249)	(0.579)	(0.098)	(0.154)
	[1.759]	[0.468]	[-0.654]	[0.826]	[5.308]	[-0.823]
$\hat{h}_{t-1}$	0.006	0.917	-0.045	-0.155	0.054	-0.017
	(0.025)	(0.040)	(0.035)	(0.082)	(0.021)	(0.048)
	[0.227]	[22.773]	[-1.294]	[-1.891]	[2.608]	[-0.357]
$\hat{c}_{t-1}$	0.058	-0.189	0.549	-0.180	-0.279	0.054
	(0.110)	(0.188)	(0.145)	(0.342)	(0.126)	(0.223)
	[0.525]	[-1.022]	[3.776]	[-0.234]	[-2.206]	[0.241]
$\hat{i}_{t-1}$	0.011	-0.013	0.075	0.612	-0.136	0.059
	(0.047)	(0.074)	(0.065)	(0.152)	(0.037)	(0.088)
	[0.235]	[-0.170]	[1.161]	[4.029]	[-3.664]	[0.674]
$\hat{x}_{t-1}$	0.127	0.264	0.169	0.486	0.195	-0.043
	(0.074)	(0.129)	(0.097)	(0.228)	(0.101)	(0.153)
	[1.715]	[2.046]	[1.746]	[2.132]	[1.929]	[-0.282]
$\hat{r}_{t-1}$	-0.170	-0.160	-0.268	-0.533	0.042	0.715
	(0.053)	(0.093)	(0.069)	(0.163)	(0.047)	(0.110)
	[-3.221]	[-1.730]	[-3.875]	[-3.274]	[0.901]	[6.527]

Note: standard deviation and t-statistics are shown respectively in ( ) and [ ]

4.2.1 Independent processes

Independent processes can be modelled on the basis of the following autoregressive matrix, whose parameters are the estimated autoregressive coefficients from the corresponding VAR equations for output (0.27) and interest rate (0.72).[19]

$$P^I = \begin{bmatrix} 0.27 & 0 & 0 \\ 0 & 0.72 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

[19] In the restricted VAR, with totally exogenous interest rate, estimates of the autoregressive coefficient of output and interest rate are 0.276 (t-value=1.789) and 0.798 (t-value=8.523).

The choice of the autoregressive coefficient for the interest rate process is straightforward, as it is given unambiguously by the estimated autoregressive coefficient in the VAR interest rate equation. The corresponding productivity coefficient is not given by the VAR. However, it can be noted from the VAR estimates that the equations of hours and investment (and by extension capital, in theory) have much more persistent processes than output. Such a fact might suggest a weaker persistence in total factor productivity. Assuming the persistence of output as a proxy for that of productivity is in some accounts a (upper-bound) compromise. Anyway, as I explore in the next section, rather than the relative persistence of the shocks, it the strength of their propagation mechanisms that matter. The propagation of interest rate shocks can be quantitatively affected by their relative persistence, but remains systematically a key qualitative feature of the model.

4.2.2 Interest rate induced processes

Now I conjecture that interest rates might affect exogenously processes of total factor productivity and/or collateral formation. Below I try to show this conjecture is empirically plausible and also consistent with the phenomena of economies facing foreign credit constraints. Different causations are associated with different choices of non-zero non-diagonal coefficients in P.

Induced collateral formation

By assuming a permanently binding credit collateral constraint on foreign net liabilities, I assert that the economy finds itself at the extreme situation in which it cannot alter, via domestic factors, such a financial condition over realistic business cycles. Long term factors, such as limited financial development, might lead the economy to such a situation, and thereafter it is sensitive to exogenous global credit shocks. Adverse innovations of this sort can transmit directly into higher interest rates and indirectly by the tightening of the country's collateral formation. The latter means the formation of collateral via capital accumulation becomes less efficient due to changes in international financial perceptions - for instance, affecting values of assets as collateral - in periods of liquidity shortage.

Such reasoning coincides with a "credit rationing" view, pioneered by Stiglitz and Weiss (1981), by which both prices and quantities respond to changes in perceptions by creditors and investors. Realistically, a rise in the country's interest rate can precede the collateral tightening. Empirically identified as exogenous, the interest rate process is a good candidate to perform the role of a leading indicator with regard to the economy's international macroeconomic and financial conditions. Therefore, adverse global credit shocks via the interest rate precede the weakening of the country's ability to form collateral . As a result, an additional exogenous dynamic is introduced between capital and liabilities, beyound that of the collateral constraint.

I set arbitrarily that a 1% rise in interest rate would cause a 2.5% fall in the efficiency to form collateral by the means of capital accumulation. The representation of the auto-regressive matrix P can be given as below:

$$P^C = \begin{bmatrix} 0.27 & 0 & 0 \\ 0 & 0.72 & 0 \\ 0 & -2.5 & 0 \end{bmatrix}$$

Induced productivity

Interest rate shocks might adversely affect total factor productivity or alternatively terms of trade. Actually, in standard SOE models terms of trade shocks tend to act similarly to productivity shocks (Kehoe and Ruhl, 2006). Such an equivalence seems even more appropriate in the case of the CCSOE model. It has been shown empirically that the deterioration of terms of trade and the worsening of the economy's international credit conditions can be correlated, acting both in a procyclical manner. Countercyclical, interest rates lead recessions and can be accompanied by real exchange rate depreciation that is behind the deterioration of terms of trade.

Recessions in the CCSOE model can be associated with periods of excessively high costs of capital. Investment is curtailed persistently by additional costs of keeping liabilities (the interest rate effect) and by a decline in the terms of trade. Sarquis (2006) shows that rather than being a source of significant shocks, the exchange rate and the terms of trade work as transmitters of interest rate shocks. Again I would argue that the interest rate is the leading measure of such processes. It Granger causes domestic variables, in particular investment, which tend to rely intensively on imports. The VAR study of Brazil indicates that, led by rises in interest rates, recessions imply not only a curtailing of investments, but also a trade balance reversal, indicating in both cases a deterioration of terms of trade.

Consistent with the choice of the auto-regressive parameter (0.27) for productivity from the output equation, I also take from the latter the estimated value for lagged interest rate (−0.17). The interest rate induced productivity representation of P is as below:

$$P^P = \begin{bmatrix} 0.27 & -0.17 & 0 \\ 0 & 0.72 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Induced productivity and collateral formation

The combination of both interest rate induced productivity and collateral formation processes can be finally represented by the auto-regressive matrix below.

$$P^{CP} = \begin{bmatrix} 0.27 & -0.17 & 0 \\ 0 & 0.72 & 0 \\ 0 & -2.5 & 0 \end{bmatrix}$$

5 Simulated Impulse Responses

The model's simulated responses are shown in Figure 2. They come from four specifications of the exogenous processes above suggested: P^I, P^C , P^P and P^{CP}. In all specifications the CCSOE model is able to produce growth persistence and (inverted) hump-shaped responses, particularly of output and consumption, following interest rate shocks. Overall such responses conform to the empirical evidence.

The mentioned properties of the propagation mechanism of the CCSOE model relies only on the implications of the credit collateral constraint. It does not depend on induced process. It is an essential feature of purely independent interest rate shocks. Moreover, these properties are not present in standard SOE models. Growth persistence would obtain from standard models only if some ad hoc elements are introduced to mimic by force such persistence, such as time to build. But even in these cases, persistence tends to be very short-lived.[20]

The independent specification P^I is a first step in matching, qualitatively, the essence of the empirical responses: inverted hump-shaped responses and trough that are sufficiently distant from the shock, indicating propagation and recession as dynamic processes. The induced processes given by the specifications P^C , P^P and P^{CP} improve quantitatively the ability of the model's simulated responses to match empirical responses. The induced collateral formation helps in aggravating the negative growth and the trough. Interest rate shocks that impair the country's collateral and productivity can bring about more dramatic implications to the economy.

In judging different specifications, a note of caution on the consumption responses should be made. Contrary to the spirit of the model, data on consumption does not exclude durables. Therefore we should not really want to account for all the excess deviation in responses of this variable.

At the same time, as with most RBC and SOE models, the CCSOE model does not generate growth persistence in investment and trade responses to match the empirical evidence, despite producing considerable (level and not growth) persistence in investment and trade responses - a feature that is absent in standard SOE models. The propagation mechanism brings enough growth persistence to the dynamics of capital capital. As in most real business cycle literature, much stronger propagating forces would be required to give investment and trade the same sort of sluggishness as consumption and output.

[20] See for instance the simulated responses in Uribe and Yue (2006), who use time to build in a standard SOE model.

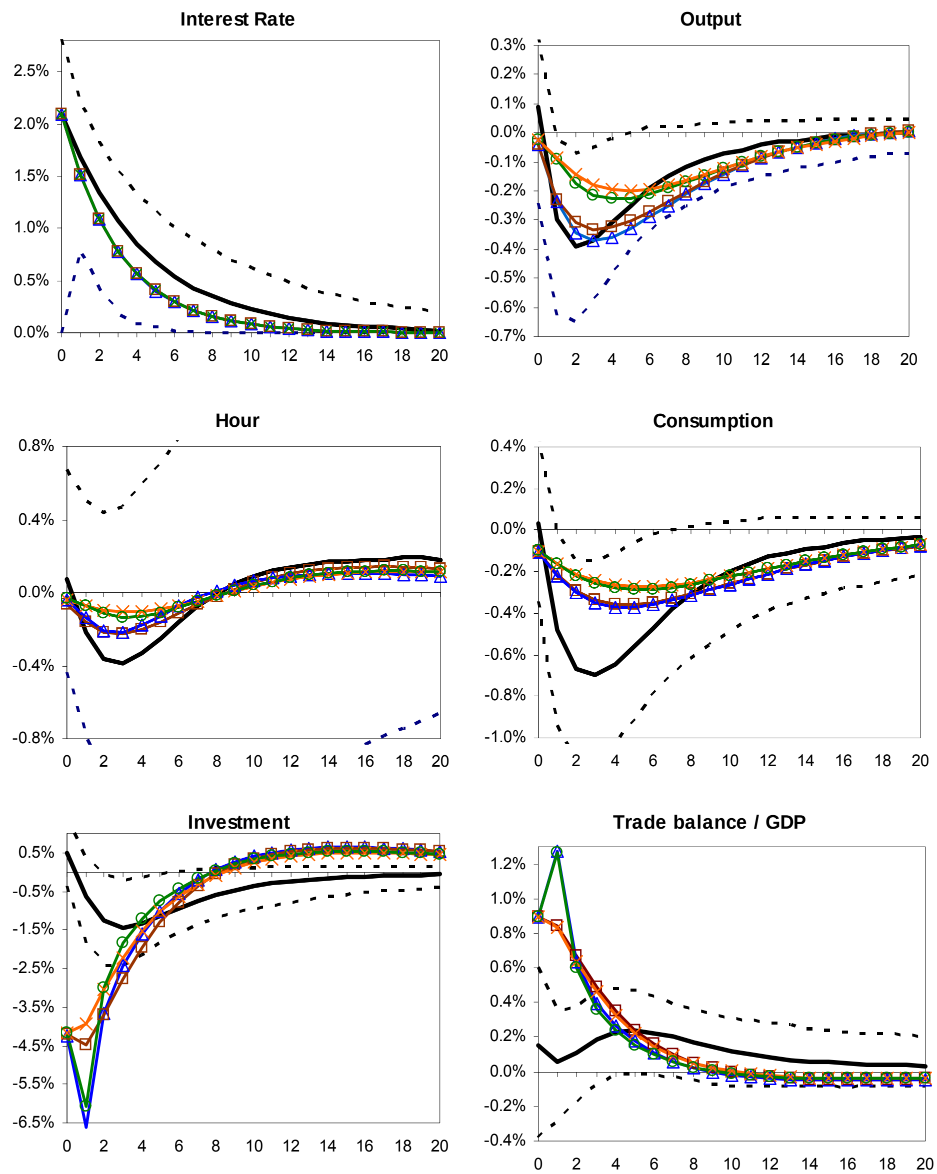


Figure 2. - Model and VAR impulse responses. Solid lines are VAR responses, accompanied by dotted lines representing 95% confidence intervals. The model responses are marked with crosses for a pure interest rate, with circles for interest rate shock and induced collateral, with squares for interest rate shock induced productivity and with triangles for interest rate shock and induced collateral and productivity. The vertical axis shows deviations from steady state. Units on the horizontal axis are quarters.

6 Sustained and Significant Effects of Interest Rate Shocks

Now I report the model's implications and results with regard to the simulated second moments. For the sake of transparency, I first report moments that obtain under a hypothetical specification in which independent productivity and interest rate processes share identical persistence and deviation of shocks. Only afterwards, I report results with the empirically motivated - independent and interest-rate induced - specifications.

The benchmark is given by Brazilian statistics (second column in Tables 5 to 7), with two caveats. First, as an alternative reference to total consumption's standard deviation, I calculate an estimate of the standard deviation of non-durable consumption, which is lower than the former.[21]. Second, I also provide along with moments of total fixed capital investment, moments of investments in machine and equipment. The latter is more volatile than the former, in part because durables are not comprehended. Such alternative statistics for consumption and investment help avoiding an automatic and sometimes misguided comparison with the simulated moments of a single-good model of RBC, which conventionally refers to nondurable consumption and to fixed capital investment, including durables consumption.

6.1 Non-neutrality

This subsection considers the non-neutrality of world interest rate shocks. I borrow this term from Mendoza (1991), meaning by it the strength by which interest rate can determine business cycle statistics, namely second moments, beyond the effects of productivity shocks. As a general property of the model, non-neutrality of interest rates is better (more strictly) assessed with the assumption of independently equivalent productivity and interest rate exogenous process. The idea behind this exercise is to be somehow agnostic on the exogenous processes, without using my priors (empirical or theoretical) about the relative importance of these processes. I choose the following identical persistence and deviation of independent shocks: $\rho^z = \rho^r = 0.78$ and $\epsilon^z = \epsilon^r = 1\%$. I analyze their implications both separately and simultaneously, so to make precise the significance of their individual effects.

The simulated moments are reported in Table 5. Interest rate shocks are responsible from 3% to 26% of the output volatility. Their relative strength is even greater in consumption and investment volatility. For instance, they can determine from 16% to 55% of consumption's variability and at least 40% of

[21] These calculations are based on domestic production series for durables and non durables.

investment's. In contrast to standard SOE models[22], non-neutrality of interest rate is assured and emerges as a central feature of the CCSOE model.

Table 5. Second moments: hypothetical specification with independent exogenous processes

	Real data	Simulated model with shocks to:		
		Productivity $\rho^{z}{}_{=0.78}$, $\rho^{r}{}_{=0}$ $\epsilon^{z}{}_{=1\%}$, $\epsilon^{r}{}_{=0\%}$	Interest Rate $\rho^{z}{}_{=0}$, $\rho^{r}{}_{=0.78}$ $\epsilon^{z}{}_{=0\%}$, $\epsilon^{r}{}_{=1\%}$	Both $\rho^{z}{}_{=0.78}$, $\rho^{r}{}_{=0.78}$ $\epsilon^{z}{}_{=1\%}$, $\epsilon^{r}{}_{=1\%}$
Standard deviation (%)				
Output	2.1	1.82	0.50	1.88
Consumption	2.7	0.91	0.60	1.09
non durables	1.8			
Investment	5.5	7.35	9.89	12.33
mach. & equip.	11.0			
Hour	3.1	1.36	0.44	1.43
Trade balance	2.3	0.38	2.01	2.05
Correlation with output				
Consumption	0.77	0.87	0.97	0.84
Investment	0.74	0.98	0.29	0.62
mach. & equip.	0.59			
Hour	0.71	0.96	0.82	0.95
Trade balance	-0.02	-0.76	-0.27	-0.21
Serial correlation				
Output	0.75	0.67	0.93	0.69
Consumption	0.69	0.85	0.92	0.87
Investment	0.67	0.69	0.66	0.67
mach. & equip.	0.70			
Hour	0.84	0.80	0.93	0.81
Trade balance	0.21	0.74	0.64	0.64

Note: The real data statistics are population moments calculated by the author based on original quarterly series for Brazil over 1990q1 to 2005q4. National accounts are seasonally adjusted. All variables are in logarithms, except trade balance ratio to GDP, and detrended with Hodrick and Prescott (1980) filter ($\lambda = 1600$). Hours only available from 1992q1. ϵ stands for standard deviation of shock.

The above ranges attributed to the impact of interest rate in real aggregates are in line with the empirical evidence. The trade balance is however excessively sensitive to interest rate. Moreover, the model can potentially reconcile the relatively higher volatility of consumption relative to output in emerging

[22] See Mendoza (1991) and Schmitt-Grohé and Uribe (2003).

economies. Interest rate shocks appear to be behind such a phenomena. Individual interest rate shocks induce 20% higher volatility in consumption than in output.

Moreover, the CCSOE model overcomes anomalies typical of standard SOE models, as the absence of serial correlation in investment and of significantly negative correlation between output and trade ratio.[23] Thanks to stronger and more persistent effects of interest rate shocks, moments become more in line with real data. Significant examples of this relate to the standard deviation of trade balance, the serial correlation of investment, and the correlations of output with investment and with trade balance.

It is worth noting that, in contrast to interest rate shocks, independent collateral formation shocks would have negligible effects in second moments over business cycles. I do not report the second moments for these shocks. With the same hypothetical persistence ($\rho^{\gamma} = 0.78$) and perturbation ($\epsilon^{\gamma} = 1\%$), they would generate on their own a negligible variation in output (of 0.03%). They would not add a significant share (less than 2%) to the overall macroeconomic volatility caused by productivity and interest rate shocks. Independent collateral formation shocks are therefore as insignificant and neutral in the CCSOE model as interest rate shocks are in standard SOE models.

The additional macroeconomic volatility of interest rate shocks indicates the strength of the model's propagation mechanism. If the deviation of the latter was twice productivity's, keeping the same persistence ($\rho^{z} = \rho^{r} = 0.78$ and $\epsilon^{z} = 1\%$, $\epsilon^{r} = 2\%$), the share of output's variability accounted by interest rate shocks would rise from the range of 3% to 27% to 12% to 47%. The range associated with consumption would rise further from 16% to 55% to 40% to 80%. The simulated standard deviation of consumption relative to output's would rise from 0.58 to 0.69.

6.2 Matching second moments

6.2.1 Independent processes

I now address independent processes of productivity and interest rates using the empirically motivated auto-regressive coefficients of P^{I}. Table 6 is equivalent to Table 5, apart from the fact that the persistence and the volatility of the shocks are empirically motivated. I am taking the benefit of my empirical priors that indicate a less persistent productivity process, with relatively weaker perturbations. As a result, interest rate shocks account for 8% to 40% of output variability and for 42% to 81% of the volatility of consumption. Note that the

[23] Also see Mendoza (1991) and Schmitt-Grohé and Uribe (2003).

empirically estimated shares of interest rate in the variance decompositions of these variables are within these simulated ranges.

Table 6. Second moments: empirically motivated specification with independent exogenous processes (P^I)

	Real data	Simulated model with shocks to:		
		Productivity $\rho^z{=}0.27$, $\rho^r{=}0$ $\epsilon^z{=}1.2\%$, $\epsilon^r{=}0\%$	Interest Rate $\rho^z{=}0$, $\rho^r{=}0.72$ $\epsilon^z{=}0\%$, $\epsilon^r{=}1.6\%$	Both $\rho^z{=}0.27$, $\rho^r{=}0.72$ $\epsilon^z{=}1.2\%$, $\epsilon^r{=}1.6\%$
Standard deviation (%)				
Output	2.1	1.58	0.69	1.73
Consumption	2.7	0.60	0.84	1.03
non durables	1.8			
Investment	5.5	6.31	14.88	16.16
mach. & equip.	11.0			
Hour	3.1	0.89	0.58	1.06
Trade balance	2.3	0.38	3.08	3.10
Correlation with output				
Consumption	0.77	0.86	0.96	0.77
Investment	0.74	0.99	0.23	0.44
mach. & equip.	0.59			
Hour	0.71	0.93	0.83	0.89
Trade balance	-0.02	-0.36	-0.21	-0.11
Serial correlation				
Output	0.75	0.26	0.92	0.36
Consumption	0.69	0.59	0.91	0.80
Investment	0.67	0.34	0.62	0.58
mach. & equip.	0.70			
Hour	0.84	0.54	0.93	0.66
Trade balance	0.21	0.45	0.59	0.59

Note: The real data statistics are population moments calculated by the author based on original quarterly series for Brazil over 1990q1 to 2005q4. National accounts are seasonally adjusted. All variables are in logarithms, except trade balance ratio to GDP, and detrended with Hodrick and Prescott (1980) filter ($\lambda = 1600$). Hours only available from 1992q1. ϵ stands for standard deviation of shock.

6.2.2 Interest rate induced processes

Interest rate induced processes of productivity and/or collateral formation further strengthen the role of interest rate shocks in the CCSOE. Table 7 displays the second moments for the three specifications suggested in the previous Section. Interest rate induced productivity and collateral can act as an additional mechanism to further transmit shocks and to augment the economy's overall volatility. They mainly help in amplifying more dramatically the effects of interest rate shocks, and they do so within the propagation and growth persistence mechanisms already in place at the core of the model, as propelled by purely independent shocks.

Table 7. Second moments: interest rate induced (collateral and/or productivity) specifications

	Real data	Simulated model with induced:		
		collateral P^C $\epsilon^z{=}1.2\%,\ \epsilon^r{=}1.6\%$	productivity P^P $\epsilon^z{=}1.2\%,\ \epsilon^r{=}1.6\%$	both P^{CP} $\epsilon^z{=}1.2\%,\ \epsilon^r{=}1.6\%$
Standard deviation (%)				
Output	2.1	1.77	1.98	2.05
Consumption	2.7	1.07	1.27	1.31
non durables	1.8			
Investment	5.5	19.46	17.67	20.96
mach. & equip.	11.0			
Hour	3.1	1.11	1.35	1.41
Trade balance	2.3	3.80	3.15	3.83
Correlation with output				
Consumption	0.77	0.78	0.82	0.83
Investment	0.74	0.35	0.57	0.49
mach. & equip.	0.59			
Hour	0.71	0.89	0.90	0.90
Trade balance	-0.02	-0.07	-0.28	-0.24
Serial correlation				
Output	0.75	0.39	0.47	0.50
Consumption	0.69	0.81	0.84	0.84
Investment	0.67	0.58	0.64	0.61
mach. & equip.	0.70			
Hour	0.84	0.68	0.74	0.76
Trade balance	0.21	0.58	0.61	0.59

Note: ϵ stands for standard deviation of shocks. In all specifications ϵ^γ=0

To ascertain which specification is the most appropriate or realistic is not simple. Table 8 gathers a summary of comparative measures of the share of output variability explained by interest rate shocks, in the presence of productivity shocks, resulting from both the independent and the induced processes. The VAR exercise (see Table 2) suggests that interest rate shocks can determine from 23% to 37% of the output variability. All proposed specifications lead to simulated shares of output variability that are consistent with the empirically estimated range.

Table 8. Real macroeconomic variability explained by interest rate shocks under different specifications of exogenous processes

Specification of exogenous processes	$St.dev.(Y)$ explained by interest rate shock (%)		
	minimum	maximum	average
Independent, P^I	8	40	24
Induced collateral, P^C	11	45	28
Induced productivity, P^P	20	59	39
Induced collateral & productivity, P^{CP}	23	63	43

Note: $St.dev.(Y)$ refers to the standard deviation of output.

Table 9 reveals the potential of different specifications of interest rate induced processes in magnifying fluctuations. The induced specifications can add up to 18% of real macroeconomic volatility. At the same time, productivity induced specifications can augment the relative variability of consumption. Overall the weaker impact of induced collateral results from the fact that it has poorer propagating properties than productivity.

Table 9. Additional volatility derived from interest rate shocks under different induced processes

Specification of exogenous processes	$\frac{St.dev.(Y)}{St.dev.(Y^I)}$	$\frac{St.\ dev.(C)}{St.\ dev.(Y)}$
Independent, P^I	1	0.60
Induced collateral, P^C	1.02	0.60
Induced productivity, P^P	1.14	0.64
Induced collateral & productivity, P^{CP}	1.18	0.64

Note: $St.dev.(Y^I)$ refers to the standard deviation output with the independent P^I specification of the exogenous process.

6.3 Serial correlation with interest rates

The four empirically motivated specifications (P^I, P^C, P^P and P^{CP}) of the exogenous processes generate simulated serial correlations between output and interest rates that match qualitatively and quantitatively the empirical correlations of unconditional data. Figure 3 below depicts these cross-correlations. As expected in an environment of sustained and significant effects of interest rate shocks over realistic business cycle horizons, the correlations over time are in agreement with the hump-shaped pattern of both estimated and simulated output responses, respectively in Figures 1 and 2.

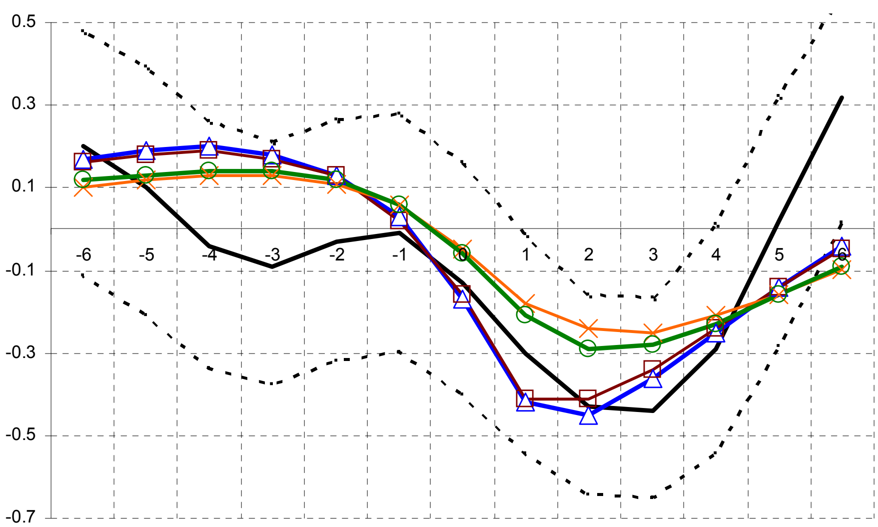


Figure 3. Serial correlations between GDP an interest rates, $\mathrm{corr}(r_{t+j}, y_t)$. Solid (black) lines represent empirical unconditional correlations, with their 95% confidence intervals shown in dotted lines. The model's simulated correlations with independent shocks (P^I) are marked with (orange) crosses. The alternative model's correlations, with interest rate induced processes are marked with (green) circles for the P^C case, with (brown) squares for the P^P case, and with (blue) triangles for the P^{CP} case. The vertical axis represent correlations and the horizontal axis j quarters.

7 Understanding the Model

The CCSOE model focuses on (international) financial and macroeconomic aspects that are absent in standard SOE models: first, net foreign liabilities are permanently constrained; and, second, the credit constraint is dynamically subject to the accumulation of capital. Let me recall that, beyond being a factor of production, capital also serves as collateral and thus has a financial value. Thus, following an adverse (temporary) interest rate shock, investment and consumption expenditures fall, reducing the economy's reliance on foreign savings. These effects are clear from the resource constraint. They are however subject to the impatience and credit constraint assumptions. The economy has an incentive to operate through gradual cuts in these expenditures, spreading the adjustment over time. Such an incentive comes precisely from the use of capital as a collateral in order to sustain foreign financing and to smooth consumption. Such a delayed adjustment imposes nevertheless further costs of financing over contractions. It results eventually in more prolonged and aggravated recessions.

Capital's dual role as collateral and factor of production imposes a premium between the marginal product of capital and interest rate, which triggers the propagation mechanism associated to more prolonged and aggravated recessions. This is implicit in the interplay of the non-standard FOCs with regard to capital and liabilities - respectively equations (10) and (11). On the one side, an additional multiplier, representing the value of collateral, appears in the Euler equation (10) governing consumption growth dynamics, as shown in Kocherlakota (2000), Arellano and Mendoza (2002) and Chari et al (2005). On the other side, the same term enters the Euler equation (11) governing capital accumulation. Deprived from adjustment costs and with $\sigma = 1$, the two equations result in the following:

$$\beta E \frac{\lambda_{t+1}}{\lambda_t} \frac{R_{t+1}}{1-\varphi_t} = 1 \tag{10'}$$

$$\beta E \frac{\lambda_{t+1}}{\lambda_t} \left[\frac{Z_{t+1} f_K(K_t, L_{t+1}) + 1 - \delta}{1 - \gamma_t \varphi_t} \right] = 1 \tag{11'}$$

By substituting out the multiplier in equations (10') and (11'), we have:

$$\beta E \frac{1}{1-\gamma_t} \frac{\lambda_{t+1}}{\lambda_t} \left[Z_{t+1} f_K(K_t, L_{t+1}) + 1 - \delta - \gamma_t R_{t+1} \right] = 1 \tag{12}$$

Through the intermediation of collateral's value in the two previous conditions, Equation (12) is a combined or extended intertemporal Euler condition. Note crucially in equation (12) that the interest rate is negatively correlated with consumption growth, in sharp contrast to the positive correlation that obtains in standard SOE models. Shocks to the interest rate lead to recessions, with negative growth persistence. The economy reacts to such shocks by a sluggish response in capital. Therefore, implicit in equation (12) lies a premium between the marginal product of capital and the interest rate. Following adverse interest rate shocks, the premium recovers from below its steady state value and later overshoots this value, just before stabilizing. Recessions and recoveries are roughly associated with premium respectively below and above its steady state level. Consumption and output responds therefore in an (inverted) hump-shaped form.

It can be shown through Equation (12) that intertemporal propagation processes can be affected directly by shocks to productivity, interest rate and collateral. A temporary shock to productivity propagates further due to the collateral parameter. Note in (12) that consumption growth is proportional to $(1-\gamma_t)^{-1}$. Moreover, this term gives the amplitude of the growth rate and therefore the pattern of adjustment following shocks. $\gamma \to 1$ would imply a prompt adjustment, while $\gamma \to 0$ would lead to a minimum rate of adjustment. Only intermediary values of γ $(0 < \gamma < 1)$ give sluggish and pronounced adjustments.

The Euler equations of the CCSOE models nest the intertemporal consumption substitution problems of standard business cycle models both of closed-economy and of SOEs, as special extreme cases. The latter obtains when the multiplier approaches zero ($\varphi_s \to 0$); that is, when collateral has no value in production. Note that in this case equations (10) and (11) conform to: $1 = \beta E \frac{\lambda_{t+1}}{\lambda_t}[Z_{t+1} f_K(K_t, L_{t+1}) + 1 - \delta]$ and $1 = \beta E \frac{\lambda_{t+1}}{\lambda_t} R_{t+1}$. Implicit in this lies the permanent equivalence (no premium) between interest rate and marginal product of capital: $R_{t+1} = Z_{t+1} f_K(K_t, L_{t+1}) + 1 - \delta$. Such equivalence is only consistent with equation (12) if the economy had the ability to use 100% of its capital as a collateral ($\gamma \to 1$). Approaching this limit, the dynamics of the CCSOE model becomes closer to that of standard SOE models. At the limit, it would be deprived of its characterizing credit friction features. On the other hand, approaching $\gamma \to 0$ leads to a dynamics convergent with RBC models, when exchanges with the world economy do not take place.

The credit frictionless conditions of standard SOE models, resulting in an automatic equivalence between interest rate and marginal product of capital, appear to be behind the "neutrality" of exogenous world real interest rate that they exhibit. In comparison to the CCSOE model, the standard models are only able to produce recoveries. Moreover, those recoveries reveal poor propagation, since they simply tend to mimic the dynamics of the exogenous interest rate process. Such lack of propagation and growth persistence is overcome by the CCSOE framework thanks to its stronger and richer inter-temporal consumption substitution and capital adjustment problems, which are born inter-twined and inter-connected by the role of capital as collateral.

Now we can understand the fact that within the credit constrained framework the rest of the world economy would behave similar to a closed-economy RBC model. Note that international assets would in equilibrium be dictated by: $B_t = -B_t^*$. Still for the simplest case (with no adjustment costs and $\sigma = 1$), it can be shown that $\gamma_t^* = \frac{-B_t^*}{K_t^*} = \gamma_t \frac{K_t}{K_t^*}$. The smaller the economy, relative to the rest of the world, the lower γ_t^*. In such a case, interest rate shocks have an increasingly negligible role.

8 Robustness

8.1 Shock persistence

In the baseline simulations, the choice of the autoregressive coefficients for both productivity and interest rate shocks is dictated by the unrestricted VAR coefficients ($\rho^z = 0.27$ and $\rho^r = 0.72$). The restricted VAR with exogenous interest rate would suggest a marginally higher ρ^r, close to 0.80, and the RBC literature usually attributes a higher value for ρ^z. Raising the autoregressive coefficients requires lower standard deviations for the shocks in order to match the overall volatility of the macroeconomic series and their serial correlation with interest rate. While higher ρ^z might weaken somehow the role of pure interest rate shocks and reduce the relative volatility of consumption in explaining business cycle fluctuations, higher ρ^r would have the opposite effects. The moments shown in Tables 5 and 6 illustrate these effects.

8.2 Collateral formation parameters

In the CCSOE collateral formation parameters are key to determining the steady state levels of consumption and investment. At the same time they can alter the dynamics of these variables. Higher values for either σ or γ strengthen the economy's capacity to take the benefit of collateral formation and therefore expose it to more amplified volatility in investment and consumption relative to output. Relative excess in consumption's volatility is a feature of the CCSOE model that is observed in emerging economies. Conditional on sharing the features of a CCSOE, emerging economies with stronger collateral formation - either by domestic merits or by international endorsement - might reveal larger volatility in consumption.

8.3 Functional form of utility

The specification of utility I use is a general one, mixing features of both GHH and KPR preferences, although closer to the latter, as $b = 0.15$. A higher (lower) value for b would diminish (augment) the overall volatility of the economy and also the relative volatility of consumption. Underlying both effects is the reduced (expanded) growth persistence of shocks. Distancing from GHH preferences and approaching KPR preferences enhances the short term responsiveness of labour. Figure 4 shows responses to interest rate shocks under the specification with induced productivity & collateral formation (P^{CP}). There I contrast the baseline response (line marked with triangles) with the alternative response under $b = 0.20$ (dotted line). Higher b exacerbates the sluggishness in the response of labour, as the latter becomes less sensitive to wealth effects. Strict GHH preferences would further prolong recessions.

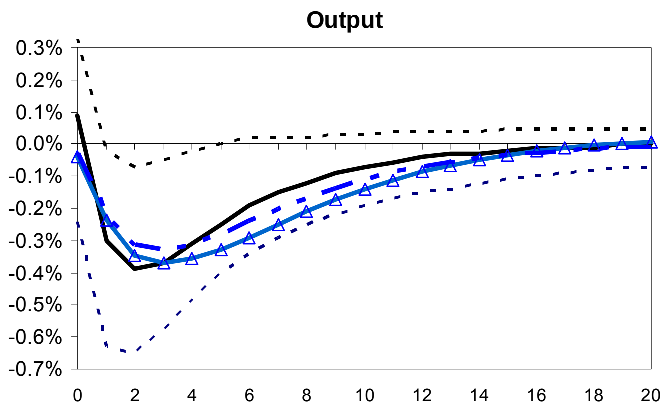


Figure 4. Effects of higher b within the induced productivity & collateral formation specification (P^{CP}). The vertical axis shows deviations from steady state. Units on the horizontal axis are quarters.

9 Conclusion

In this paper I highlight the main empirical regularities that characterize the role of the (exogenous) world real interest rate in determining business cycles of Brazil, as an example of an emerging economy, and, consistently, develop a credit constrained small open economy (CCSOE) model which matches comprehensively most of these regularities.

I use VAR analyses of Brazil - a representative, and important, case study among emerging economies - to illustrate that real interest rates, which are countercyclical and lead cycles, work not only as key transmitters of exogenous global credit shocks, but also as major drivers of real macroeconomic fluctuations. Real interest rate shocks help explain the excessive volatility of output and the higher relative volatility of consumption in emerging economies. Moreover, estimated responses of output and consumption, as well as of investment and trade to GDP, to interest rate shocks, are hump-shaped. They denote the existence of strong propagation mechanisms which are able to prolong and aggravate recessions.

In contrast to standard SOE models, including extensions with working capital constraints, the model I propose exhibits considerable propagation of interest rate shocks. Their effects are sustained and significant over business cycle frequencies, augmenting macroeconomic volatility, beyond the effects of shocks to productivity - or equivalently to the terms of trade. The model successfully produces hump-shaped responses of output and consumption. The latter variable is more volatile than output, potentially magnifying consumption's relative variability in settings with a greater variety of shock sources. The model also does a good job in replicating second moments, in a more realistic way than usually achieved by standard approaches. It matches systematically the correlations between output and interest rates at different lags, particularly over recessions. Furthermore, it manages to avoid some of the recurrent anomalies of SOE models with respect to investment and trade moments.

The model relies fundamentally on an impatience assumption and a permanently binding endogenous credit collateral constraint on foreign liabilities. It stresses the role of international credit frictions, which, by the means of collateral, provide foreign creditors and investors with some control or discipline over the country's international financial exposure.[24] It also emphasizes some reliance on foreign financing, perhaps due to insufficient financial deepening (relative to economic development) in some emerging economies.

The essence of the model resides in the inter-play of two non-standard intertemporal Euler equations, governing capital dynamics and consumption

[24] Such frictions can be seen somehow as complementary to and/or consistent with the home bias puzzle in international debt and equities.

growth. These equations provide considerable growth persistence to the propagation mechanism of interest rate shocks. The economy responds to such adverse (temporary) shocks by curtailing investment and consumption in accordance with persistent declines in collateral. Collateral formation results from the accumulation of capital, which, besides serving as a factor of production, has a key role in securing sustained foreign financing and consumption smoothing. Subject to such incentives, accruing from the dual role of capital, the economy reacts to shocks on a sluggish and persistent basis, which end up prolonging and aggravating recessions.

The paper shows consistently that international credit constraints can make an economy prone to suffer from disturbances in international financial markets that can be transmitted via shocks to world interest rates. In doing so it shares some of the concerns manifested in Aghion, Bachetta and Banerje (2004), Aoki, Benigno and Kiyotaki (2006), Caballero and Krishnamurthy (2001) and Gertler, Gilchrist and Natalucci (2003). It also contributes in exploring further, in a simple and quantitatively tractable model, the macroeconomic implications of financial frictions at the level of SOEs, in the spirit of Bernanke and Gertler (1989), Carlstrom and Fuerst (1997), and Kiyotaki and Moore (1997). Following the business cycles tradition, the paper treats cycles as transitory fluctuations around a stable growth path and support the understanding that emerging economies' excessive volatility and other business cycle features derive not necessarily from domestic factors, but rather from international financial sources. It contrasts with the view espoused for instance by Aguiar and Gopinath (2007) that cycles are derived from shocks to trend growth in these economies.

Future research might seek to integrate explicitly an endogenous country-spread within the credit constrained small open economy (CCSOE) framework. I do not differentiate shocks to country-spread from innovations to international interest rates. There is the scope to explore the feedback effects from the domestic economy to the interest rate, via the country spread, as argued among others by Arellano (2005). Another challenge ahead is to research further the additional propagation that can potentially arise from the leading effects of interest rate on other factors that help determining the economy's international financial and macroeconomic conditions, such as collateral and terms of trade. In this paper, such effects are only addressed exogenously. Mechanisms by which their additional propagation would obtain endogenously merit further investigation.

A Appendix

All Brazilian quarterly series are available at IPEAData (www.ipeadata.gov.br), the on-line macroeconomic database of the *Instituto de Pesquisa Pura e Aplicada* - IPEA. The national account series are originally from the *Instituto Brasileiro de Geografia e Estatística* (IBGE) - www.ibge.gov.br. Further details of the data are the following:

Output. Real GDP from the IBGE National Accounts. The series was expressed in natural logarithmic of available seasonally adjusted series.

Hours. Industrial hours from the *Confederação Nacional das Industrias* (CNI). The series was expressed in natural logarithmic of available seasonally adjusted index.

Consumption. Total consumption from the IBGE National Accounts. The series was expressed in natural logarithmic of the available seasonally adjusted index.

Durables. Production of consumption durable good from the IBGE National Accounts. The series was expressed in natural logarithmic of the available seasonally adjusted index.

Investment. Total fixed capital investment from the IBGE National Accounts. The series was expressed in natural logarithmic of the available seasonally adjusted index.

Machine and equipment. Total investment in machine and equipment from the IBGE National Accounts. The series was expressed in natural logarithmic of the available seasonally adjusted index.

Ratio of export to import. The series was calculated by the author as a ratio of net export to GDP. Official foreign trade statistics are produced by the Ministry for Development, Industry and Foreign Trade - Ministério do Desenvolvimento, da Indústria e do Comércio Exterior (MDIC).

Country spread. EMBI JP Morgan Plus index for Brazil. The series is available from Datastream.

US inflation and 3-month nominal interest rate are available from the FRED database of the Federal Reserve Bank of St Louis.

References

[1] Agénor, Pierre-Richard, McDermott, C. John & Prasad, Eswar S. (2000). Macroeconomic Fluctuations in Developing Countries: some stylized facts. The World Bank Economic Review, 14(2), 251-85.

[2] Aghion, Philippe, Bachetta, Philippe & Banerjee, Abhijit (2004). Financial Development and the Instability of Open Economies. Journal of Monetary Economics, 51, 1077-1106.

[3] Aguiar, Mark & Gopinath, Gita (2007). Emerging Market Business Cycles: the cycle is the trend?. Journal of Political Economy, 115(1), 69-102.

[4] Aoki, Kosuke, Benigno, Gianluca & Kiyotaki, Nobuhiro (2006). Adjusting to Capital Account Liberalization. mimeo, London School of Economics.

[5] Arellano, Cristina (2005). Default Risk and Income Fluctuations in Emerging Economies. mimeo, University of Minnesota.

[6] Arellano, Cristina & Mendoza Enrique (2002). Credit Frictions and 'Sudden Stops' in Small Open Economies:An Equilibrium Business Cycle Framework for Emerging Markets Crises. NBER Working Paper No. 8880.

[7] Backus, David K., Kehoe, Patrick J. & Kydland, Finn E. (1992).International Real Business Cycles. Journal of Political Economy, 100(4), 745-775.

[8] Becker, Gary S. & Mulligan, Casey B. (1997). The Endogenous Determination of Time Preference. The Quarterly Journal of Economics, 112(3), 729-58.

[9] Bernanke, Ben & Gertler, Mark. (1989). Agency Costs, Net Worth, and Business Fluctuations. American Economic Review, 79, 14-31.

[10] Blanchard, Olivier (1985). Debt, Deficits, and Finite Horizons, Journal of Political Economy, 93 (2), 223-47.

[11] Blankenau, William, Kose, M. Ayhan, & Yi, Kei-Mu (2001). Can World Real Interest Rates Explain Business Cycles in a Small Open Economy? Journal of Economic Dynamics & Control, 25, 867-889.

[12] Caballero, Ricardo & Krishnamurthy, Arvind (2001). International and Domestic Collateral Constraints in a Model of Emerging Market Crises. Journal of Monetary Economics, 48, 513-548.

[13] Calvo, Guillermo, Leiderman, Leonardo & Reinhart, Carmen (1996). Inflows of Capital to Developing Countries in the 1990s. The Journal of Economic Perspectives, 10(2), 123-39.

[14] Campbell, John & Deaton, Angus (1989). Why is Consumption so Smooth? Review of Economic Studies, 56(3), 357-374.

[15] Canova, Fabio (2005). The transmission of US shocks to Latin America. Journal of Applied Econometrics, 20, 229-251.

[16] Carsltrom, Charles T. & Fuerst, Timothy S. (1997). Agency Costs, Net Worth and Business Fluctuations: a computable general equilibrium analysis. American Economic Review, 87(5), 893-910.

[17] Caselli, Francesco & Feyrer, James (2007). The Marginal Product of Capital. Quarterly Journal of Economics, 122(2), 535-568.

[18] Charri, V. V. Kehoe, Patick J. & McGrattan, Ellen R. (2005). Business Cycle Accounting. Federal Reserve Bank of Minneapolis Research Department Staff Report 328.

[19] Charri, V. V. Kehoe, Patick J. & McGrattan, Ellen R. (2005). Sudden Stops and Output Drops. American Economic Review, 95(2), 381-387.

[20] Chrisitiano, Lawrence J., Eichenbaum, Martin & Evans, Charles. (2005). Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy. Journal of Political Economy, 113(1), 1-45.

[21] Cogley, Thimothy & Nason, James N. (1995). Output Dynamics in Real-Business-Cycle Models. American Economic Review, 85(3), 492-511.

[22] Ellery Jr, Roberto, Gomes, Victor & Sachsida, Adolfo. (2002). Business Cycle Fluctuations in Brazil. Revista Brasileira de Economia, 56(2), 269-308.

[23] Gertler, Mark, Gilchrist, Simon & Natalucci, Fabio (2003). External Constraints on Monetary Policy and the Financial Accelerator. NBER Working Paper No. 10128.

[24] Gollin, Douglas (2002). Getting Income Shares Right. Journal of Political Economy, 110(2), 458-474.

[25] Greenwood, Jeremy, Hercowitz, Zvi, & Huffman, Gregory W. (1988). Investment, Capacity Utilization, and the Real Business Cycle. American Economic Review, 78(3), 402-417.

[26] Huh, Hyeon-Seung (2005). A Simple Test of Exogeneity for Recursively Structured VAR models. Applied Economics, 37, 2307-13.

[27] Issler, Joao V. & Piqueira, Natalia S. (2001). Estimating Relative Risk Aversion, the Discount Rate, and the Intertemporal Elasticity of Substituion in Consumption for Brazil Using Three Types of Utility Function. Brazilian Review of Econometrics, 20(2), 201-239.

[28] Jaimovich, Nir & Rebelo, Sergio (2006). Can News about the Future Drive the Business Cycles? NBER Working Paper No. 12537.

[29] Kanczuk, Fabio (2004). Real Interest Rate and Brazilian Business Cycles. Review of Economic Dynamics, 7(2), 436-455.

[30] Kehoe, Timothy J. & Ruhl, Kim J. (2007). Are Shocks to the Terms of Trade, Shocks to Productivity? NBER Working Paper No. 13111.

[31] King, Robert. & Rebelo, Sergio (1999). Resuscitating Real Business Cycles, in J. Taylor and M. Woodford (eds.), Handbook of Macroeconomics, North-Holland, 927-1007

[32] King, Robert, Plosser, Charles I. & Rebelo, Sergio T. (1988). Production, Growth, and Business Cycles: I. The Basic Neoclassical Model. Journal of Monetary Economics, 21(2/3), 195-232.

[33] Kiyotaki, Nobuhiro & Moore, John. M. (1997). Credit Cycles. Journal of Political Economy, 105(2), 211-248.

[34] Kiyotaki, Nobuhiro (1998). Credit and Business Cycles. Japanese Economic Review, 49, 18-35.

[35] Kocherlakota, Narayana R. (2000). Creating Business Cycles Through Credit Constraints. Federal Reserve Bank of Mineapolis Quarterly Review, 24(3), 2-10.

[36] Kose, M. Ayhan, Prasad, Eswar, Rogoff, Kenneth & Wei, Shang-Jin (2006). Financial Globalization: a reappraisal. NBER Working Paper No. 12484.

[37] Kydland, Finn E. & Zarazaga, Carlos E. J. M. (1997). Is the Business Cycle of Argentina "Different"? Federal Reserve Bank of Dallas Economic Review, 4th Quarter, 21-36.

[38] Mendoza, Enrique (1991). Small Open Economy Real Business Cycle. American Economic Review, 81(4), 797-818.

[39] Mendoza, Enrique (2006). Endogenous Sudden Stops in a Business Cycle Model with Collateral Constraints: a Fisherian dDeflation if Tobin's Q. NBER Working Paper No. 12564.

[40] Neumeyer, Pablo A. & Perri, Fabrizio (2006). Business Cycles in Emerging Markets: the role of interest rates. Journal of Monetary Economics, 52(1), 345-380.

[41] Oviedo, Pedro M. (2005). World Interest Rate, Business Cycles, and Financial Intermediation in Small Open Economies. mimeo, Iowa State University.

[42] Paasche, Bernhard (2001). Credit Constraints and International Financial Crises. Journal of Monetary Economics, 48(3), 623-650.

[43] Sarquis, Sarquis J B. (2006). Global Credit Shocks, Interest Rate and Business Cycles in Emerging Economies. mimeo, London School of Economics.

[44] Sarquis, Sarquis J B. (2006). Business Cycles in a Credit Constrained Small Open Economy. mimeo, London School of Economics.

[45] Schmitt-Grohé, S. & Uribe, M. (2003). Closing Small Open Economy Models. Journal of International Economics, 61, 163-185.

[46] Stiglitz, Joseph & Weiss, Andrew (1981). Credit Rationing in Markets with Imperfect Information. American Economic Review, 71, 393-410.

[47] Stock, James & Watson, Mark (1999). Business Cycle Fluctuations in U.S. Macroeconomic Time Series, in J. Taylor and M. Woodford (eds), Handbook of Macroeconomics, North-Holland, 1, 3-64

[48] Uhlig, Harald (1999). A Toolkit for Analysing Nonlinear Dynamic Stochastic Models Easily, in R. Marimon and A. Scott (eds), Computational Methods for the Study of Dynamic Economies, Oxford, 30-61.

[49] Uribe, Martín & Yue, Vivian Z. (2006). Country Spreads and Emerging Countries: who drives whom? Journal of International Economics, 69(1), 6-36.

[50] Val, Paulo R. da C. & Ferreira, Pedro C. (2001). Modelos de Ciclos de Negócios Aplicados à Economia Brasileira. Pesquisa e Planejamento Econômico, 31(2), 213-248.

[51] Yaari, Menahem E. (1965) Uncertain Lifetime, Life Insurance, and the Theory of the Consumer. Review of Economic Studies, 32(2), 137-150.

Regionality Revisited: An Examination of the Direction of Spread of Currency Crises

Amil Dasgupta Roberto Leon-Gonzalez Anja Shortland

This draft: June 2006

Revised for the Journal of International Money and Finance

Abstract: What determines the direction of spread of currency crises? We examine data on waves of currency crises in 1992, 1994, 1997, and 1998 to evaluate several hypotheses on the determinants of contagion. We simultaneously consider trade competition, financial links, and institutional similarity to the "ground-zero" country as potential drivers of contagion. To overcome data limitations and account for model uncertainty, we utilize Bayesian methodologies hitherto unused in the empirical literature on contagion. In particular, we use the Bayesian averaging of binary models which allows us to take into account the uncertainty regarding the appropriate set of regressors.

We find that institutional similarity to the ground-zero country, as measured by quality-of-governance indicators, plays an important role in determining the direction of contagion in all the emerging market currency crises in our dataset. We thus provide persuasive evidence in favor of the "wake up call" hypothesis for financial contagion. Trade and financial links may also play a role in determining the direction of contagion, but their importance varies amongst the crisis periods and may be sensitive to the specification of the prior.

JEL Classification: F31, F32, C11

Keywords: Financial contagion; Currency crises; Governance; Bayesian Model Averaging

Addresses for correspondence: Dasgupta (*corresponding author*) Department of Accounting and Finance, London School of Economics, London, WC2A 2AE, UK. E-mail: a.dasgupta@lse.ac.uk Leon-Gonzalez and Shortland: Department of Economics, University of Leicester, Leicester, LE1 7RH, UK

Acknowledgements: We thank the editor, James Lothian, and two anonymous referees for valuable comments that have shaped this revision. We are grateful to Hugo Benitez-Silva, Giancarlo Corsetti, Emilio Gómez-Déniz, Fernando Fernández-Rodríguez, Andrew Patton, Jorge Pérez-Rodríguez, John Rust, and Francisco J. Vázquez-Polo for helpful comments, and to Beatrice Weder for sharing the data on the 1998 crisis. Dasgupta thanks the ESRC for financial support via grant RES-156-25-0026.

1 Introduction

Currency crises tend to occur in waves. In repeated instances from the early 1970s to the late 1990s it has been observed that when speculative attacks lead to a currency crisis in one country, market volatility tends to spread to other countries in the region and elsewhere. Several mechanisms have been proposed to explain this phenomenon, generally referred to as contagion. Commonly discussed mechanisms include the transmission of crises through trade and financial links between countries, as well as the (rational) updating of beliefs by financial traders about the sustainability of specific institutional and developmental models. The latter is sometimes referred to as the "wake-up call" theory of financial contagion.

In this paper we empirically evaluate the relative importance of a number of potential transmission mechanisms that have been proposed in the existing literature, by analysing four waves of currency crises in the 1990s. We make two contributions.

First, we simultaneously include institutional (quality-of-governance) variables alongside the trade, finance and macroeconomic variables commonly analysed in empirical literature on contagious currency crises, thereby directly testing the "wake-up call" hypothesis.

Second, we utilize Bayesian methodologies hitherto unused in the empirical literature on contagion to overcome model uncertainty and data limitations. In particular, we use Bayesian averaging of binary models, which allows us to take into account the uncertainty regarding the set of regressors that should be included in the empirical analysis of contagion.

Before proceeding further, it is worth clarifying the remit of our exercise. In this paper we do not seek to enter the debate on whether contagion exists. While there are now several theoretical equilibrium models of contagion, there is not yet complete empirical agreement about whether contagion exists.[1] In this paper, we simply *assume* that contagion exists and aim only to shed light on the mechanisms by which it may propagate.

[1] See Dungey and Tambakis (2003) for a discussion of the term "contagion" as well as Dungey *et al* (2003) for a detailed review of the contagion literature.

Much of the extant empirical literature on contagious currency crises stresses the phenomenon of regional contagion. It focuses on trade and financial links, which tend to occur in geographical clusters, and finds evidence in favor of both as potential transmission mechanisms for contagion.[2] However, the currency crises of the 1990s have spread far beyond the region of the original crisis country. Glick and Rose (1999) deem that Hong Kong, Indonesia, the Philippines and Thailand were affected by the "Mexican crisis" in 1994/1995, while Argentina, Brazil, the Czech Republic, Hungary and South Africa are considered to have been among the victims of the Asian Crisis. According to Van Rijckeghem and Weder (2001) the Russian crisis of 1998 affected 16 countries outside the former Soviet Union, including Argentina, Hong Kong, Indonesia, South Africa and Turkey. While trade competition in third markets or financial links may be possible explanations for extra-regional contagion, it is also interesting to examine the possibility that a speculative attack on a country follows from a "wake-up call" regarding a specific model of development: a currency crisis in one country may highlight vulnerabilities associated with a particular set of institutional features, which may also be found in other countries outside the region.

There is now much data measuring the institutional features of different countries. Our paper contributes to the literature by directly testing the extent to which institutional similarity with the "ground zero" country determines the direction of spread of currency crises. This is done while simultaneously considering standard factors such as trade competition and financial links to give an overall view of the drivers of financial contagion in foreign exchange markets.

In addition, our paper utilizes recent econometric methodology that is relevant to the empirical analysis of financial contagion. There is no universally agreed-upon theoretical model of contagion: several alternative hypotheses coexist. In the presence of such model uncertainty, Bayesian model averaging (BMA) is a natural candidate for empirical work in this area. The idea of BMA was first proposed by Leamer (1978). It is a tool for forecasting and estimation when the researcher does not know the true model. Starting from a prior where all possible models are considered to be equally good, the

---

[2] Eichengreen *et* al (1996), Glick and Rose (1999), Kaminsky and Reinhart (2000), Caramazza et al (2000), Van Rijckeghem and Weder (2003).

method allows researchers to estimate the posterior probabilities of the models, using the data, and then weight their estimates and forecasts from each model by such posterior probabilities. While BMA has recently been extensively used in applied problems (see various references below), we are the first to use it in the context of financial contagion.

In addition to this, Bayesian methods allow us to overcome data limitations. Empirical samples in the contagion literature are of necessity small: in all previous studies the number of observations is below 100 countries. Of these only a small subset experience a crisis in each episode of contagion. Unlike maximum likelihood, Bayesian methods are also valid in small samples.

Summary of Results

We examine data on currency crises in 1992, 1994, 1997, and 1998, focussing on the relative importance of trade, financial links, and institutional similarity on the direction of contagion. We report the following results:

1. Institutional (quality-of-governance) variables play a vital role in the spread of all emerging market currency crises in our dataset. Following a crisis in the "ground zero" country, countries that are, ceteris paribus, institutionally similar have a higher probability of experiencing a currency crisis. In the crises of 1994, 1997, and 1998, the increase in crisis probabilities due to institutional similarity ranges between 24% and 63%. Our results, therefore, provide substantial empirical support for the "wake-up call" hypothesis for financial contagion.

   In contrast, however, institutional similarity has less explanatory power in the 1992 EMU crisis, confirming the intuition that the "wake-up call" hypothesis is most relevant for emerging markets.
2. Other factors, such as financial links (through common lenders) and trade also play a role in determining the direction of contagion, but their importance may vary across crisis periods. For example, financial links appear to be important in the 1998 crisis, while trade competition is important for 1997. The relevance of these variables is also sensitive to the specification of the prior.

Our paper is linked to a large and growing literature on financial contagion. In what follows, we briefly survey this literature.

2 Literature review

The literature has considered a number of potential channels for international financial contagion.[3] The first potential channel derives from international trade.[4] If a country experiences a sharp devaluation it gains a competitive advantage over its trade partners and over competitors in third markets. To the extent that (the expectation of) deteriorating current account deficits signals potential currency weakness, countries with strong trade connections to the "ground zero country" become more likely to experience a speculative attack. Glick and Rose (1999) examine the importance of the trade channel and find statistical evidence from cross-country data that currency crises spread among countries which have strong trade links.

A second potential channel of contagion derives from financial linkages between countries.[5] Here contagion arises because groups of countries rely on common creditors and investors. If a country experiences a speculative attack, its major creditor banks may experience liquidity problems, which undermine their ability to provide emergency finance to other countries or trigger capital outflows to restore capital adequacy ratios. Therefore, countries which rely on external funding from the same creditors and investors as the "ground zero country" become vulnerable to speculative attacks. The importance of the "common creditor effect", meaning contagion through bank lending, has been empirically examined by Van Rijckeghem and Weder (2001 and 2003), Caramazza *et al*. (1999), Hernandez and Valdes (2001) and Kaminsky and Reinhart (2000). The results indicate that vulnerability to speculative attacks can spread among clusters of countries which depend on the same lenders. Caramazza *et al*. (1999) additionally show that countries which are more important to the common lenders are more likely to become crises countries than those which only receive a very small proportion of the common lenders' total lending.

---

[3] See Pericoli and Sbracia (2001) and Dungey *et al* (2003) for literature reviews. Chui *et al* (2004) sets out the framework for assessing external vulnerabilities in more detail.

[4] For a theoretical formalization of this idea see, for example, Gerlach and Smets (1995).

[5] For theoretical models formalizing this hypothesis, see, for example, Goldstein and Pauzner (2005), Allen and Gale (2000), and Dasgupta (2004).

A third channel for contagion derives from shared updating by market participants about the sustainability of specific institutional frameworks or development models. Such a view of contagion is commonly referred to as the "wake-up call" hypothesis.[6] The argument here is that if a country with a particular development strategy, institutional set-up or macroeconomic situation experiences a devaluation, this may be seen as revealing information about the vulnerability of countries of a similar "type" and hence cause the spread of crises.[7] A good example of a major re-evaluation of an economic development strategy was seen in the rapid turn-around in 1997 from applauding the "Asian Miracle"[8] to deploring the "Asian Debacle". Months before the crisis South East Asia's "dedicated capitalism"[9] and "Asian values" were praised and held up as strategies for successful development the world over, but were swiftly condemned as "crony capitalism" in the immediate aftermath of the crisis and held responsible for economic vulnerabilities. Issues such as "corruption", "regulatory quality" and "transparency" suddenly came to the forefront of investor attention and may have contributed to the spreading of the crisis to countries perceived to have similar deficits in accountability and data quality. While a large literature has emerged in recent years to measure and quantify the effects of legal and institutional variables on financial development[10] and financial fragility[11] to our knowledge no direct test of the impact of institutional similarity on financial contagion has been carried out. It is a contribution of this paper to provide a direct examination of the "wake-up call" hypothesis using measures of institutional similarity provided in the literature.

---

[6] The term "wake-up call" originates from Goldstein (1998). For theoretical formalizations of this hypothesis, see Rigobon (1998) and Basu (1998). Van Rijckeghem and Weder (2003) provide evidence for the "wake-up call" hypothesis from the Russian crisis, which caused generalized outflows from emerging markets.

[7] See Drazen (1998) on "information externalities"

[8] See for example the 1993 World Bank publication "*The South East Asian Miracle*" hailing the "fundamentally sound development policies" and "tailored government interventions" in eight high performing Asian economies.

[9] Porter (1996)

[10] See Beck and Levine (2003) for a review

[11] Demirgüç-Kunt and Detragiache. (1998), Kaminsky and Reinhart (1999), Kaminsky and Schmukler (2003)

3 Data

In Table 1 we summarize the variables that we use. For a given wave of currency crises and for each country *i*, the dependent (binary) variable records whether country *i* experienced a currency crisis following the crisis in the ground zero country.[12] It is taken from Glick and Rose (1999) for 1992, 1994 and 1997 and from Van Rijckeghem and Weder (2001) for 1998.

To quantify the trade channel for contagion we use the "trade share" indicator computed by Glick and Rose (1999) for 1992, 1994 and 1997 and Van Rijckeghem and Weder (2001) for 1998. A high value of this index indicates that the country's exports compete intensely with the ground zero country in third markets.

To measure financial links between countries, we choose two indicators of competition for funds based on Caramazza *et al*. (1999). Define the "common lender" to be the creditor country most exposed to the ground zero country. For any given country, our first indicator indexes the importance of the common lender to that country. For the emerging market crises the "common lenders" are the US (1994), Japan (1997) and Germany (1998). For example, in the Russian crisis of 1998 the indicator looks at the proportion of country i's total borrowing which derived from German banks. Our second indicator measures how important a potential target country is to the common lender. Thus, the indicator measures country i's borrowing as a proportion of the total loans made by the common lender. We also include a multiplicative interaction of these two indicators. The data are taken from the Bank for International Settlements' (BIS) consolidated data, covering bank lending from banking systems in the "reporting area" of 18 industrialised countries to countries outside the "reporting area".[13] All indicators refer to banks' position reported at the date closest to the respective crises i.e. December 1994 for the Mexican crisis, June 1997 for the Asian crisis and June 1998 for the Russian crisis. The BIS data only cover lending from the reporting area to countries outside the reporting area, meaning that no financial data are available for the 1992 crisis in the

[12] Glick and Rose use journalistic and academic histories of crises episodes to identify countries suffering from contagion, Van Rijckeghem and Weder (2001) utilise a panel of IMF experts.

[13] The reporting area countries are: Austria, Belgium, Canada, Denmark, Finland, France, Germany, Ireland, Italy, Japan, Luxembourg, Netherlands, Norway, Spain, Sweden, Switzerland, UK and the US.

European exchange rate mechanism. However, it is likely that contagion through financial centres is a phenomenon limited to emerging market currency crises.

Our analysis of the "wake-up call" hypothesis is based on a number of variables that have been used in the literature to capture institutional similarity between countries. To measure institutional similarity, we use a number of variables taken from the set of governance indicators compiled by Kaufman *et al*. (1999) for the World Bank. In particular, we test whether variables such as corruption, regulatory quality, and the degree to which the rule of law is upheld influence whether investors withdraw capital from a country. A disadvantage of this dataset is that data collection only began in 1996. However, Kaufmann *et al.* (2005) note that the quality of governance tends to be highly persistent, because institutions change only slowly.[14] Changes in governance over time are small relative to the level of the governance indicators and the reported error margin on the estimates. Changes in annual governance estimates where the 90% confidence intervals do not overlap are only reported in a tiny minority of countries.[15] We therefore take the average score of each country in the years 1996, 1998 and 2000 and used this for each episode of the 1990s currency crises. For each country, and for each relevant variable, we then compute a measure of *similarity* to the ground zero country. For example, let c_i be the corruption index for country *i* that is constructed as just described, and let c_0 be the same variable measured for the ground zero country. Then the variable that we use in our analysis is defined as: $|c_i - c_0|/|c_o|$. An analogous index of similarity is constructed for the other two institutional variables.

An additional way of capturing institutional similarity derives from legal origin. The large literature on law and finance (e.g. La Porta *et al*. 1998) argues that a country's legal system (mostly acquired through colonisation or occupation) has important effects on how confidently investors transact in a country, and that this differs significantly between Anglo-Saxon common law and French, German and Scandinavian civil law systems.[16] Motivated by this literature, we complement our core measures of institutional similarity summarized above by an indicator of common legal origin, which takes the

[14] http://www.worldbank.org/wbi/governance/pdf/GovMatters_IV_main.pdf
[15] http://www.worldbank.org/wbi/governance/pdf/govmatters3_wber.pdf
[16] See Beck *et al* (2001) for a review

value 1 if a country has the same legal system as the ground zero country. The data are taken from La Porta *et al*. (1998).

We include relative geographical distance to the ground zero country as a "control variable" in our regressions. Relative distance is relevant as a control for at least two reasons. First, trade competition and financial links tend to be regionally clustered, and thus it is worth considering these effects after controlling for pure geographic regionality. Second, countries that are closer are likely to have more similar institutions and culture. Thus, relative distances may also capture institutional similarity not captured by the more direct measures above. The distances between countries were computed as the distances between capital cities, using the distance calculator provided by Darrell Kindred[17] at http://www.indo.com/distance.

Finally, we use a number of macro-economic variables as additional control variables, such as current account and budget deficits, countries' reserve positions, credit expansion, inflation and growth performance. These variables control for the possibility that a country would have fallen into crisis regardless of the attack on the first country, because of its own weak macroeconomic fundamentals.[18] In our choice of control variables, we have been guided by the prior work of Eichengreen *et al*. (1996), Glick and Rose (1999) and Van Rijckeghem and Weder (2003). The variables are computed or taken from the IFS for the period preceding the crisis[19]. This reflects both the delay in data becoming available and the fact that in the immediate aftermath of a currency crisis there is usually a significant worsening of the macroeconomic situation.

4 Methodology

4.1 Bayesian Model Averaging

Let Z be the $n \times k$ matrix that contains all variables that could potentially enter in the regression equation, where n is the number of observations and k is the number of potential regressors. Let $Y = (y_1, ..., y_n)'$ be an $nx1$ vector of observed binary variables.

[17] This calculator uses the latitudes and longitudes of the cities concerned and then computes the distance between them by using the Geod program, which is part of the PROJ system, a set of cartographic projection tools, provided by the US Geological Survey at ftp://kai.er.usgs.gov/pub/.
[18] See e.g. Kaminsky *et al* (1998) for a review of the empirical currency crises literature
[19] 1994 for Mexico, 1996 for Asia and 1997 for Russia

We consider all binary probit models that result from including a different subset of Z as explanatory variables. This gives rise to 2^k models. In particular, model M_j is defined as the following probit model:

$$Y^* \mid \theta \sim N\left(Z_j^{'}\theta_j, I_n\right), \quad y_i = \begin{cases} 1 & if \quad y_i^* \geq 0 \\ 0 & if \quad y_i^* < 0 \end{cases}$$

where $Y^* = (y_1^*,...,y_n^*)'$ is an $nx1$ vector containing unobserved latent data, Z_j is a $n \times k_j$ submatrix of Z, θ is a $k \times 1$ vector of unknown parameters, θ_j is a $k_j \times 1$ subvector of θ containing the elements of θ that are included (i.e., not restricted to be zero) in model M_j, and I_n is the identity matrix of dimension n.

Our inference for θ is based on the posterior mean and credible regions[20] of the posterior density of θ ($\pi(\theta \mid Y, Z)$), which is a weighted average of the posterior densities obtained under each of the models ($\pi(\theta \mid Y, Z, M_j)$):

$$\pi(\theta \mid Y, Z) = \sum_{j=1}^{2^k} \pi(M_j \mid Y, Z) \quad \pi(\theta \mid Y, Z, M_j)$$

Here, $\pi(M_j \mid Y, Z)$ represents the posterior probability of model M_j, which is given by Bayes' Rule as follows:

$$\pi(M_j \mid Y, Z) = \frac{\pi(M_j) \quad \pi(Y \mid Z, M_j)}{\sum_{l=1}^{2^k} \pi(M_l) \quad \pi(Y \mid Z, M_l)} \text{ with } \pi(Y \mid Z, M_j) = \int \pi(Y \mid \theta, Z, M_j) \pi(\theta \mid M_j) d\theta$$

where $\pi(M_j)$ is the prior probability of model j, $\pi(\theta \mid M_j)$ is the prior density of θ under model M_j, and $\pi(Y \mid \theta, Z, M_j)$ is the likelihood.

We now define a crucial concept. The *probability of inclusion* for a (possibly singleton) set of explanatory variables S_j is the joint posterior probability of all models that include at least one of the variables in S_j. In other words, the probability of inclusion of S_j is the probability that at least one variable in S_j has a non-zero effect on the expected outcome of the dependent variable. Thus, a zero inclusion probability implies that all of the coefficients in θ that correspond to S_j are equal to zero. Inclusions probabilities will

[20] A 95% credible interval is the Bayesian analogue of a frequentist 95% confidence interval, and it is an interval that contains the true value of the parameter with probability 95% (e.g. see Koop 2003, p. 44).

be crucial to interpreting our results: variables with high posterior inclusion probabilities are relevant determinants of contagion; others are not.

Our Bayesian methodology presents two important advantages over its more commonly used classical counterparts in the context of the contagion literature. First, as we have already noted, it allows us to control for model uncertainty. Second, Bayesian methods are valid in small samples. Both of these properties make Bayesian methods particularly suitable for the empirical analysis of financial contagion.

4.2 Prior

We use a prior that is computationally convenient and relatively uninformative. For each model M_j, we choose a normal prior as follows:

$$\theta_j \mid M_j \sim N(0, V), \qquad V_j = g(Z_j' Z_j)^{-1},\ g > 0 \tag{1}$$

This class of priors has been extensively used for Bayesian estimation (e.g. Zellner, 1986, Poirier 1985, Fernandez Ley and Steel, 2001). A prior mean of zero implies that we consider outcomes y_i=1 and y_i=0 to be equally likely *a priori* for i=1,..,n. It also implies that *a priori* each covariate is equally likely to have a positive or a negative effect. The prior variance-covariance matrix depends on the scalar parameter g. It is instructive to think of our choice of g in terms of the implied distribution of the following quantity:

$$\bar{\pi} = \Pr(y = 1 \mid \bar{z}_j, \theta_j, M_j),$$

i.e, the ex ante probability, under model M_j, that the *average* country (a country with average values of regressors) experiences a currency crisis.

While it may be tempting to make our prior "more uninformative" by choosing a very large value of g, it is easy to see that this does not necessarily result in a reasonable prior. Very large values of g imply that, *a priori*, we expect $\bar{\pi}$ to be either 1 or 0 and consequently marginal effects (on probabilities) to be approximately zero.[21] Therefore, instead of arbitrarily fixing a very large value for g, we carefully adapt priors that have been proposed in the existing literature for other related models. In particular, we use three values for g. Details of the prior-elicitation process for g are provided in Appendix A. We summarize our choices here.

[21] We comment further on this issue in Appendix A.

Our first choice for g is given by:

$$g = \overline{g} = \left(\overline{z}_j^{'}\left(Z_j^{'}Z_j\right)^{-1}\overline{z}_j\right)^{-1} \quad (2)$$

This choice is tantamount to assuming that the prior distribution of $\overline{\pi}$ is uniform, a choice recommended by Geisser (1984) for the estimation of a probability.

Our second choice of g is given by

$$g = 2.46\overline{g}$$

This amounts to assuming that the *a priori* distribution of $\overline{\pi}$ is approximately Beta(½,½), a prior recommended in the literature for the estimation of probabilities (Lee 1987). Compared to the uniform prior, the Beta prior gives slightly more weight to values of $\overline{\pi}$ near to 0 and 1. Finally, for sensitivity analysis we also consider $g = 5\overline{g}$.

We carry out our computations for all three values of g.

4.3 Computation

For our computations, we use the algorithm of Holmes and Held (2006) who extend the methodology of Albert and Chib (1993) to allow for model uncertainty. The Holmes and Held algorithm is a Markov Chain that visits a model (M_n) at each iteration n, and also generates a value for θ conditioning on M_n and the data. A priori all models are given equal probability. Starting with any arbitrary initial model and starting value of θ, Holmes and Held (2006) show that, as the number of iterations increases, the models and parameter values generated can be regarded as a sample from the true posterior distribution of models and parameters. Therefore, posterior means and other quantities of interest can be easily approximated with their sample analogues. The posterior probability of model M_j is given by the proportion of iterations that visit model M_j. We provide details of the algorithm in Appendix A.

The results below derive from 165,000 iterations of the algorithm. The first 5,000 iterations are discarded. Essentially identical results were obtained with an independent run of fewer (65,000) iterations, indicating good convergence.

5 Results

Our main economic results are presented in Tables 2 and 3. Tables 4 and 5 assess the out-of-sample predictive power of the models. The dependent variable is binary, taking value one if the country concerned suffered a crisis. For each independent variable we report three quantities. First, we report the probability of inclusion of the variable (*p*), as defined in Section 4.1. This is the probability that the effect associated with a regressor is different from zero. Second, since Probit coefficients are hard to interpret, we report the posterior mean for the marginal effect of each variable. These marginal effects are evaluated at the sample mean of variables.[22] Third, for each marginal effect, we include the 95% credible interval, as defined in Section 4.1. This is the Bayesian analogue to the classical 95% confidence interval in a Maximum Likelihood estimation. Finally, at the bottom of each table, we report the *joint* inclusion probability for the institutional similarity variables (*R. Law*, *Reg. Q.* and *Corrupt*) and for the finance variables (*Fi*1, *Fi*2 and *Fi*1**Fi*2). Since our goal is to understand whether trade competition, financial links, or institutional similarity drive financial contagion, it is important for us to compare the joint probabilities of inclusion of these different categories of variables. The results reported in Tables 2 and 3 correspond to the prior with $g = \bar{g}$ 2.46. The results that we comment upon are robust to the 3 choices of g, unless otherwise stated.[23]

Institutions

The main conclusion from our empirical analysis is that institutional similarity is an important predictor of financial contagion during the emerging market crises of 1994, 1997, and 1998. With our two core priors, the joint probability of inclusion of the institutional similarity variables is at least 94% in all crises episodes with the exception of 1992.[24] In 1992 the joint probability is above 80%, which is high but not conclusive. For the emerging market crises of 1994, 1997 and 1998, credible intervals at 95% for the

[22] Note that since we have a dummy variable among the regressors, namely Legal Origin, by taking the sample mean of variables we are evaluating the marginal effect at the average intercept. The marginal effect for the dummy variable Legal Origin is calculated as the change in probability when Legal Origin changes from 1 to 0. The marginal effects for the finance variables (Fi1 and Fi2) take into account the consequent change in the interaction variable Fi1*Fi2.

[23] Results with the other two priors are available at http://fmg.lse.ac.uk/~amil/research.html.

[24] When the prior has $g = 5\bar{g}$, the joint probability of institutions in 1998 is still high but decreases to 90%.

marginal effects of institutional variables almost always exclude positive values, which is consistent with the wake-up call theory: countries that are institutionally similar to the ground zero country are more likely to experience crises. The only exceptions for these crises are the 95% credible intervals for *R. Law* and *Reg. Q.* in 1998, which contain positive values. However, the effects of these two variables in 1998 are more likely to be zero, since their inclusion probabilities are only 27% and 20%, respectively.[25]

Since it is difficult to interpret the size of the marginal effects of the institutional similarity variables, we now provide an alternative way of assessing whether the estimated effects are large or small. Consider a country *A* that has average value for all regressors except for the institutional similarity variables (*R. Law*, *Reg. Q.* and *Corrupt*), all of which take value 0: i.e. the country is identical to the ground zero country with respect to institutions. In addition, consider a country *B* that also has average value for all regressors, but whose institutional variables take the same value as the country in our sample that is the most dissimilar, in terms of institutions, to the ground zero country[26]. Hence, countries *A* and *B* are different only with respect to institutions. Then, country *A* is affected by the crisis in years 1994, 1997, 1998 with probabilities (24%, 63%, 55%), whereas the corresponding probabilities for country *B* are zero for each year.[27] This confirms that institutional similarity played a particularly important role in the direction of spread of the emerging market crises of 1994, 1997 and 1998.

Our results on the effects of common legal origin are less emphatic. Zero values can never be confidently ruled out for the effect of *Legal Or* in any of the crises, especially in 1992, in which the effect of this variable seems to be negligible. The probability of inclusion of *Legal Or*. is highest in year 1997, in which positive values can be ruled out, indicating that countries with the same legal system as the ground zero country experienced lower probability of crisis. The 1997 ground zero country has British legal origin, which suggests that overall countries with British legal origin were *ceteris*

[25] Even for the EMU crisis of 1992, where institutional effects are clearly less important, the 95% credible interval for *Reg. Q.*, the only variable with substantial inclusion probability, excludes positive values.

[26] The most dissimilar country in our sample is defined as the country that maximises the Euclidean distance with respect to the ground zero country. In terms of the variables that are defined in Section 3 and Table 1, it maximises $(R.\ Law)^2 + (Reg.\ Q)^2 + (Corrupt)^2$. According to our data, the most dissimilar countries to the ground zero countries (in terms of institutions) for 1992 (Finland), 1994 (Mexico), 1997 (Thailand) and 1998 (Russia) were Guinea-Bissau, Singapore, New Zealand and Singapore, respectively.

[27] For 1992 this probability decreases from 8% to zero.

paribus less susceptible to financial crises, which is consistent with the results of the Law and Finance literature.[28] The opposite effect is observed in years 1994 and 1998, where the ground zero countries have French and (Post-Socialist) civil law legal origins respectively. However, in these years the probability of the effect being zero is high.

We now turn to the other potential channels for financial contagion. Our results suggest that, after controlling for institutional similarity, other variables such as financial linkage, trade competition and distance have limited impact. We provide a detailed discussion in what follows.

Finance

The joint probability of inclusion of finance variables is above 90% only for the 1998 crisis, provided that the prior variance g is equal to $\bar{g}$ or 2.46 $\bar{g}$. This probability decreases to 84% when $g = 5\bar{g}$. Despite the high joint probability, the individual inclusion probabilities of $Fi1$, $Fi2$ and $Fi1*Fi2$ are low. This is probably caused by multicollinearity. Despite of the problem of multicollinearity, it can be observed that the marginal effects of $Fi1$ and $Fi2$ in 1998 are positive, since credible intervals exclude negative values. Furthermore, the size of the mean marginal effects is non-negligible. Although the effect is not as clear for other years, the evidence for 1998 confirms the intuition that the more dependent the country is on the common lender, the more likely it is that it will be affected by the crisis.

Trade and Distance

The inclusion probability of *Trade* is highest in the 1997 crisis. It is 94% when $g = 5\bar{g}$, but it is below 90% when g is equal to $\bar{g}$ or 2.46$\bar{g}$ (and it then takes values 81% and 87%, respectively). However, 95% credible intervals indicate that the possibility of negative values can be confidently neglected, and that mean marginal effects are sizeable, indicating therefore that the trade channel of contagion was probably important in 1997. Although zero values are more likely in 1994 and 1998, 95% credible intervals indicate that moderately large effects are still possible, and negative values are very unlikely.

[28] See Beck *et al* (2001)

Therefore, *Trade* could also have been an important determinant in years 1994 and 1998. However, in contrast with Glick and Rose (1999), we find that *Trade* has a negligible effect in the 1992 crisis. It is *Distance*, instead, that seems to play an important role. *Distance* in 1992 is probably simply capturing the fact that EMU countries, which happen to be geographically near, were much more likely to be affected by the crisis. However, the negligible effect of trade is not caused by accounting for distance: if we excluded distance from the set of *potential* regressors the marginal effect of trade would continue to be small.

Out of Sample Predictions

We evaluate the predictive performance of the model using the prior with $g = \bar{g}$ 2.46 and the following predictive rule, which is defined for p= 0.5, 0.65, 0.75, 0.9, 0.95:

- y_i is predicted to be one when the posterior mean of $\Pr(y_i = 1|Z) > p$.
- y_i is predicted to be zero when the posterior mean of $\Pr(y_i = 1|Z) < 1-p$.

Predictions are made for (1997, 1998) based on parameter estimates from 1994 data. Similarly, predictions are made for (1994, 1998) based on parameters estimated with 1997 data, and for (1994, 1997) based on 1998 data. For each of these three cases we calculate two error rates: E_0 is the proportion of observations that were predicted to be zero but were actually 1. Similarly, E_1 is the proportion of observations that were predicted to be one but were actually 0. Tables 4 and 5 show the results.

Table 4 shows that there are very few countries for which the posterior probability of a crisis is high, and this introduces a small sample bias in our estimate of E_1. For example, if 1994 data is used to predict the 1997-98 crises, four cases have a posterior probability of a crisis greater than 0.90. Three of these, Indonesia, the Republic of Korea and Malaysia in 1997 actually suffered a crisis.[29] Given the small number of cases that are predicted to be 1, the estimate of the error rate is bound to be imprecise.

Table 5 shows that E_0 is equal to 0 when 1-p is 0.05, and it is smaller than 0.1 when 1-p is 0.1. This suggests that the model produces reliable predictions of zeros, in

[29] A crisis is also (incorrectly) predicted for China in 1997.

the sense that a small posterior mean of $\Pr(y_i = 1|Z)$ can be taken as strong evidence against the occurrence of a crisis.

6 Conclusions

We contribute to the empirical literature on financial contagion by considering institutional similarity to the ground-zero country, measured via governance indicators, as a determinant of the direction of spread of currency crises. We find that for the emerging market crises of 1994, 1997, and 1998, institutional similarity played a substantial role in determining the direction of contagion. Simultaneously, we consider more traditional channels of contagion, including trade and financial links. We are thus able to establish the relative importance of these various channels.

Our analysis also utilizes recent econometric methodology that is relevant to the analysis of financial contagion. In the absence of a single unified model of financial contagion, researchers are faced with model uncertainty in estimation and prediction. We use Bayesian model averaging to overcome these problems, a method hitherto unused in the literature on financial contagion.

Our results provide direction to theoretical modelers on the right mix of ingredients that should go into a potential unified model of financial contagion.

Bibliography

Akaike, H, (1978), "A new look at the Bayes Procedure," *Biometrika*, 65, 53-59.

Albert, J.H. and Chib, S. (1995), "Bayesian Analysis of Binary and Polychotomous Response Data," *Journal of the American Statistical Association* 88, 669-679.

Allen, F. and Gale, D. (2000), "Financial Contagion," *Journal of Political Economy*, 108, 1-33

Basu R. (1998): Contagion Crises: The Investors' Logic, mimeograph UCLA Los Angeles

Beck, T., Demirgüç-Kunt, A. and Levine, R.E. (2001), "Legal Theories of Financial Development" *Oxford Review of Economic Policy*, 17, 483-501

Beck, T., and Levine, R.E. (2005), "Legal Institutions and Financial Development" in C. Menard and M. Shirley, (eds.), *Handbook of New Institutional Economics*, Kluwer Dordrecht (The Netherlands).

Berger, J. and Selke, T. (1987), "Testing of a point null hypothesis: the irreconcilability of significance levels and evidence (with discussion)," *Journal of the American Statistical Association*, 82, 112-139.

Bernardo, J.M. (1979), "Reference Posterior Distributions for Bayesian Inference" (with discussion), *Journal of the Royal Statistical Society*, B, 41, 113-147.

Berndt, E.R. and Savin, N.E., (1977), “Conflict among criteria for testing hypotheses in the multivariate linear-regression model,” *Econometrica*, 45, 1263-1277.

Box, G.E.P. and Tiao, G.C. (1992), *Bayesian Inference in Statistical Analysis*. Wiley (New York).

Caramazza, F., Luca, A. and Salgado, R. (2000), “Trade and Financial Contagion in Financial Crises,” *IMF Working Paper*, WP/00/55.

Dasgupta, A. (2004), “Financial Contagion through Capital Connections: A Model of the Origin and Spread of Bank Panics," *Journal of the European Economic Association*, 2, 1049-1084.

Demirgüç-Kunt, A. and Detragiache, E. (1998), “Financial Liberalisation and Financial Fragility” in *Annual World Bank Conference on Development Economics 1998*, B. Plezkovic and J Stiglitz, (eds.), 303-31.

Drazen, A. (1998), “Political Contagion in Currency Crises,” mimeo, University of Maryland.

Dungey, M, Fry, R, González-Hermosillo, B. and Martin, V.L. (2003), “Empirical Modelling of Contagion: A Review of Methodologies,” *CERF working paper*, # 8, available at: http://rspas.anu.edu.au/economics/staff/dungey/pdfs/reviewdfgm.pdf

Dungey, M. and Tambakis, D. (2003), “International Financial Contagion: What do we know?” *CERF Working Paper* # 9.

Eichengreen, B, Rose, A. and Wyplosz, C. (1996), “Contagious Currency Crises,” *NBER Working Paper* #5681.

Fernández, C, Ley, E. and Steel, M. (2001) “Benchmark priors for Bayesian model averaging,” *Journal of Econometrics*, 100, 381-427.

Geisser, S. (1984) “On prior distributions for binary trials (with discussion)” *American Statistician*, 38, 244-251.

Gerlach, S. and Smets, F. (1995): Contagious Speculative Attacks; *European Journal of Political Economy*, 11, 5-63.

Glick, R. and Rose, A. (1999), “Contagion and Trade: Why Are Currency Crises Regional?” *Journal of International Money and Finance*, 18, 603-617

Goldstein M. (1998): The Asian Financial Crisis; Institute for International Economics Washington DC

Goldstein, I. and Pauzner, A. (2004), “Contagion of Self-Fulfilling Currency Crises Due to the Diversification of Investment Portfolios,” *Journal of Economic Theory*, 119, 159-183.

Greene, W.H. (2003), *Econometric Analysis*. 5^{th} Ed., Prentice Hall (New Jersey).

Hernandez, L. and Valdes, R. (2001), “What Drives Contagion: Trade, Neighbourhood or Financial Links?” *IMF Working Paper* WP 01/29.

Holmes, C.C. and Held, L. (2006), “Bayesian Auxiliary Variable Models for Binary and Multinomial Regression,” *Bayesian Analysis,* 1, 145-168.

Jeffreys, J. (1961), *Theory of Probability*. Oxford University Press (Oxford).

Kaminsky, G., Lizondo, S. and Reinhart, C. (1998), “Leading Indicators of Currency Crises” *IMF Staff papers*, 45, 1-48.

Kaminsky, G. and Reinhart, C. (1999), “The Twin Crises: the Causes of Banking and Balance of Payments Problems,” *American Economic Review* 89, 473-500.

Kaminsky, G. and Reinhart, C. (2000), “On Crises, Contagion and Confusion,” *Journal of International Economics*, 145-168.

Kaminsky, G. and Schmukler, S. (2003), “Short-Run Pain, Long Run Gain: The Effects of Financial Liberalisation” *IMF Working Paper* 2003 / 34.

Kaufmann, D., Kraay, A. and Zoido-Lobaton, P. (1999), “Governance Matters,” *World Bank Working Paper* # 2196

Koop, G. (2003) *Bayesian Econometrics*. Wiley (London).

La Porta, R., Lopez-de-Silanes, F., Shleifer, A. and Vishny, **R.** (1998), “Law and Finance,” *Journal of Political Economy*, 106, 1113 –1155.

Lee, P. (1987), *Bayesian Statistics: An Introduction*, 1st ed., Oxford University Press (New York).

Porter, M.E. (1996), "Capital Choice: National Systems of Investment," in *As If the Future Mattered: Translating Social and Economic Theory into Human Behavior*, N.A. Goodwin (ed.), Ann Arbor: University of Michigan Press, 15-44.

Rigobon, R. (1998), “Informational Speculative Attacks: Good News is No News," *MIT Working Paper*

Selke, T., Bayarri, M.J. and Berger, J. (2001), “Calibration of P-values for testing precise null hypotheses,” *The American Statistician* 55, 62-71.

Van Rijckeghem, C. and Weder, B.S. (2001). "Sources of Contagion; Finance or Trade?" *Journal of International Economics,* 45, 293-308.

Van Rijckeghem, C. and Weder, B.S. (2003). "Spillovers through banking centers: a panel data analysis of bank flows," *Journal of International Money and Finance*, 22, 483-509.

Y	Indicator of whether country *i* experienced a currency crisis; Glick and Rose (1999) and Van Rijckeghem and Weder (2001)
Trade	Trade competitiveness as defined in Glick and Rose (1999)
Dom. Cred.	Growth of Domestic Credit
Bud/GDP	Budget Position as a percentage of GDP
CA/GDP	Current account position as a percentage of GDP
Growth	Real rate of GDP per capita growth
M2/Res	Ratio of M2 to central bank foreign reserves
Inflation	Domestic CPI inflation
GDP	GDP per capita at the beginning of the year measured in 1990 US $
Distance	Great circle distance between capitals of country *i* and ground zero country in km
Legal Or.	Legal Origin Dummy: 1 if a country has the same legal system as the ground zero country
R. Law	Similarity, to ground zero country, in the degree to which the rule of law is upheld. Decreasing with similarity. Original data from Kaufmann *et al.* (1999).
Reg. Q.	Similarity, to ground zero country, in Regulatory quality. Decreasing with similarity. Original data from Kaufmann *et al*. (1999).
Corrupt	Similarity, to ground zero country, in Levels of Corruption. Decreasing with similarity. Original data from Kaufmann *et al*. (1999).
*Fi*1	The proportion of a country's total borrowing that was borrowed from the common lender.
*Fi*2	A country's borrowing as a proportion of the total loans made by the common lender.
*Fi*1**Fi*2	The product of Fi1 times Fi2.

Table 1: Definition of variables.

	Crises in 1992				Crises in 1994			
	p	Mean and 95% credible interval for marginal effects			*p*	Mean and 95% credible interval for marginal effects		
		Lower limit	Mean	Upper limit		Lower limit	Mean	Upper limit
Trade	0.29	0	0.013	0.089	0.62	0	0.090	0.491
Dom. Cred.	0.34	0	6.0E-05	4.00E-04	0.24	-1.2E-04	-1.1E-05	8.3E-05
Bud/GDP	0.40	-0.008	-0.001	0	0.49	0	0.001	0.008
CA/GDP	0.33	-7.1E-05	3.3E-04	0.003	0.23	-1.7E-04	1.3E-04	8.3E-04
Growth	0.26	-0.003	-4.2E-04	0	0.23	-2.1E-04	1.0E-04	0.001
M2/Res	0.24	0	8.5E-05	0.001	0.35	-5.1E-04	-7.7E-05	3.1E-06
Inflation	0.46	-0.002	-3.2E-04	0	0.24	-2.5E-05	3.3E-05	2.2E-04
GDP	0.21	-4.9E-07	-2.7E-08	3.7E-07	0.32	-5.2E-06	-6.4E-07	3.9E-07
Distance	**0.92**	-3.9E-05	-1.1E-05	0	0.23	-1.7E-07	1.1E-07	1.4E-06
Legal Or.	0.12	-0.001	-4.9E-04	0	0.56	0	0.013	0.173
R. Law	0.28	-0.049	-0.002	0.033	0.55	-0.049	-0.009	0
Reg. Q.	0.62	-0.221	-0.046	0	0.37	-0.040	-0.006	0
Corrupt	0.24	-0.042	-0.002	0.025	0.47	-0.038	-0.006	0
*Fi*1	-	0	-	-	0.23	-0.057	-0.005	2.7E-02
*Fi*2	-	0	-	-	0.58	0	0.891	4.29
*Fi*1**Fi*2	-	-	-	-	0.28	-	-	-
Constant	0.42	-	-	-	0.61	-	-	-
	P(*Finance*) = - P(*Institutions*) = 0.82				P(*Finance*) = 0.78 P(*Institutions*) = **0.95**			

Table 2: Probabilities of inclusion, posterior mean and credible intervals for the crises in 1992 and 1994. *p* is the probability of inclusion of each variable. P(*Finance*) is the joint probability of inclusion of *Fi*1, *Fi*2 and *Fi*1**Fi*2. P(*Institutions*) is the joint probability of inclusion of *R. Law*, *Reg. Q.* and *Corrupt*. Prior with $g = \overline{g}$ 2.46

	Crises in 1997				Crises in 1998			
	P	Mean and 95% credible interval for marginal effects			*P*	Mean and 95% credible interval for marginal effects		
		Lower limit	Pos. Mean	Upper limit		Lower limit	Pos. Mean	Upper limit
Trade	0.87	0	0.484	1.25	0.27	-0.101	0.127	1.02
Dom. Cred.	0.40	-0.006	-1.1E-03	0	0.13	-0.004	-2.4E-04	0.001
Bud/GDP	0.64	0	0.012	0.044	0.18	0	0.003	0.027
CA/GDP	0.38	-0.013	-0.002	0	0.17	0	0.002	0.015
Growth	0.34	-0.019	-2.2E-03	0.002	0.18	0	0.004	0.036
M2/Res	**0.95**	0	0.011	3.0E-02	0.24	0	0.005	0.032
Inflation	0.22	-1.1E-03	1.1E-05	0.001	0.14	-0.002	2.8E-04	0.004
GDP	0.24	-1.1E-05	-1.1E-06	1.1E-06	0.20	-3.5E-06	3.1E-06	3.5E-05
Distance	0.24	-9.6E-06	-8.6E-07	1.8E-06	0.29	0	7.8E-06	5.1E-05
Legal Or.	0.74	-0.324	-1.3E-01	0	0.34	0	0.159	0.838
R. Law	0.83	-0.230	-0.081	0	0.27	-0.213	0.021	0.358
Reg. Q.	0.43	-0.285	-0.055	0	0.20	-0.020	0.020	0.212
Corrupt	0.21	-0.023	-0.002	8.0E-04	0.82	-0.681	-0.289	0
Fi1	0.22	-0.184	0.098	1.11	0.17	0	0.433	1.22
Fi2	0.21	-0.948	0.122	1.72	0.35	0	6.53	12.33
*Fi1*Fi2*	0.21	-	-	-	0.59	-	-	-
Constant	0.44	-	-	-	0.49	-	-	-
	P(*Finance*) = 0.51 P(*Institutions*) = **0.98**				P(*Finance*) = **0.90** P(*Institutions*) = **0.94**			

Table 3: Probabilities of inclusion, posterior mean and credible intervals for the crises in 1997 and 1998. *p* is the probability of inclusion of each variable. P(*Finance*) is the joint probability of inclusion of *Fi*1, *Fi*2 and *Fi*1**Fi*2. P(*Institutions*) is the joint probability of inclusion of *R. Law*, *Reg. Q.* and *Corrupt*. Prior with $g = \overline{g}$ 2.46

	p	0.5	0.65	0.75	0.90	0.95
1994	E_1	0.56	0.38	0.30	0.25	0.00
	NP	39	16	10	4	1
	AN	27	27	27	27	27
1997	E_1	0.71	0.69	0.68	0.67	0.00
	NP	45	36	22	6	1
	AN	23	23	23	23	23
1998	E_1	0.56	0.29	0.20	n.a.	n.a.
	NP	16	7	5	0	0
	AN	18	18	18	18	18

Table 4: Out of Sample Predictions of 1. y_i is predicted to be one when the posterior mean of $\Pr(y_i = 1|Z) > p$. When the models are estimated with 1994 data, predictions are made for (1997, 1998). Similarly, predictions are made for (1994, 1998) based on 1997 data, and for (1994, 1997) based on 1998 data. *NP* is the number of observations predicted to be 1. E_1 is the proportion of *NP* that was actually 0. *AN* is the actual number of ones in the validation sample.

	1-*p*	0.5	0.35	0.25	0.1	0.05
1994	E_0	0.15	0.09	0.09	0.04	0.00
	NP	68	54	45	25	12
	AN	80	80	80	80	80
1997	E_0	0.15	0.14	0.12	0.06	0.00
	NP	67	59	52	34	17
	AN	89	89	89	89	89
1998	E_0	0.12	0.08	0.06	0.09	0.00
	NP	89	75	54	32	10
	AN	87	87	87	87	87

Table 5: Out of Sample Predictions of 0. y_i is predicted to be zero when the posterior mean of $\Pr(y_i = 1|Z) < 1-p$. When the models are estimated with 1994 data, predictions are made for (1997, 1998). Similarly, predictions are made for (1994, 1998) based on 1997 data, and for (1994, 1997) based on 1998 data. *NP* is the number of observations predicted to be 0. E_0 is the proportion of *NP* that was actually 1. *AN* is the actual number of zeros in the validation sample.

Appendix A: Details of Bayesian Methodology

A.1: Prior Elicitation for the parameter g.

We comment first on why we do not simply choose a very large value for *g*. It is easy to see that choosing very high values for *g* (which results in a very high prior variance) results in priors that put all probability weight on *y=0* or *y=1*. For example, suppose that there is only one regressor in the model and no constant term. A sufficiently large prior variance for the slope coefficient implies that the probability that $Z\theta$ is in the interval (-4,4) is approximately zero. Note that in order to predict the outcome of y_i it does not matter in practice if $Z\theta$ is –5 or –250, since both values result in the probability of y_i=1 being approximately equal to zero. Therefore, since a large prior variance effectively rules out that $Z\theta$ lies in (-4,4), the size of the slope coefficient is no longer relevant, and all we would need, should the prior information be true, in order to predict *perfectly* the outcome of y_i, is the sign of the slope coefficient. Thus, because the prior would be so informative, the only relevant information that we would expect from the data would concern the sign of the slope coefficient. A large amount of data would be necessary to change such strong prior beliefs on large probabilities and small marginal effects.

We comment next on the three values of *g* that we actually choose. Our first choice of prior fixes a value of *g* such that:

$$Var(\bar{z}_j'\theta_j) = \bar{z}_j V_j \bar{z}_j' = 1$$

where $\bar{z}_j$ is a $k_j\times 1$ vector containing the average sample values of Z_j. This implies the following value of *g*:

$$\bar{g} = \left(\bar{z}_j'\left(Z_j'Z_j\right)^{-1}\bar{z}_j\right)^{-1}$$

To see why this choice is appealing, recall that $\bar{\pi} = \Pr(y=1 \mid \bar{z}_j, \theta_j, M_j) = \Phi\left(\bar{z}_j'\theta_j\right)$, where Φ is the distribution function of a standard normal and therefore $\bar{\pi}$ is the probability of (y=1) for a country with average values for the regressors. If we fix *g* to be equal to $\bar{g}$, then our prior for $\bar{\pi}$ is a uniform in the interval (0,1).[30][31]

[30] To see why, note that using the second fundamental theorem of calculus, the Jacobian from $\bar{\pi}$ to $\tilde{z} = \bar{z}_j'\theta_j$ is the density function of a standard normal evaluated at $\tilde{z}$.

[31] In addition, we note that if Z contains an intercept term, then expression (2) is equal to *n*. A value of *g* equal to *n* has been recommended in the context of model selection in linear models by Fernandez, Ley and Steel (2001).

Another popular choice of non-informative prior to estimate a probability is a Beta(1/2,1/2). In the context of a binomial likelihood, this prior is uninformative according to alternative criteria used by different authors (Jeffreys, 1961, Box and Tiao, 1973**,** Akaike, 1978 and Bernardo, 1979). Compared to the uniform prior, the Beta prior gives slightly more weight to values near to zero and near to 1. In our model, this implies that values of θ_j that were further away from zero would receive greater prior weight. Within our framework, we can achieve this greater weight by choosing $g = a\overline{g}$, with a>1. After experimenting with several values for a, we found that a=2.46 results in a prior for $\overline{\pi}$ that approximates well to a Beta (1/2,1/2). This is illustrated in Figure 1, which shows that our prior for $\overline{\pi}$ when a=2.46 is virtually undistinguishable from the Beta prior. Therefore, the second prior that we consider results from fixing $g = 2.46\overline{g}$. Finally, for sensitivity analysis we also consider prior (1) with $g = 5\overline{g}$.

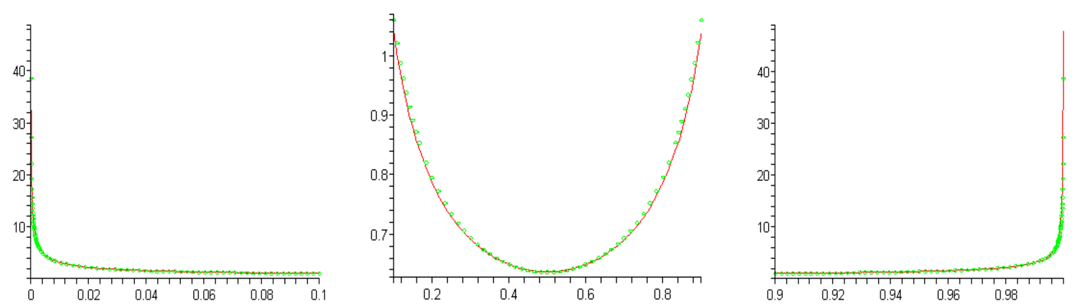


Figure 1: Three views of our prior density for $\overline{\pi}$ with $g = 2.46\overline{g}$ (continuous line) and a Beta(1/2,1/2) (dotted line).

A.2: Computation.

Let M_n be the model visited in the n^{th} iteration of the Markov Chain algorithm, let θ_n be the value of the non-zero parameters in M_n at the n^{th} iteration and similarly let Y_n^* be the value of Y^*. Assuming prior (1) for the parameters in a model, and assuming that all possible models have equal prior probabilities, the iteration (n+1) proceeds as follows:

1) Choose a model M^* from a uniform distribution defined on the following set of models:
   - Model M_n
   - Models that result from dropping one regressor in M_n
   - Models that result from adding one regressor to M_n

2) Set M_{n+1} equal to M^* with probability:

$$\alpha = \min\left\{1, \frac{(1+g)^{-k*/2}\exp\left(-1/2\left(Y_n^*\right)'\left(I_n - 1/(1+g)Z_* V_* Z_*'\right)Y_n^*\right)}{(1+g)^{-k_n/2}\exp\left(-1/2\left(Y_n^*\right)'\left(I_n - 1/(1+g)Z_n V_n Z_n'\right)Y_n^*\right)}\right\}$$

where I_n is the identity matrix of dimension n, Z_* is a $n \times k_*$ matrix with the set of regressors contained in M^*, k_n is the number of regressors in M_n and V_* and V_n are defined as in (1). Set M_{n+1} equal to M_n with probability 1-α.

3) Draw θ_{n+1} from a normal density with covariance matrix ($\tilde{V}$) and mean ($\tilde{\mu}$) equal to:

$$\tilde{V} = \frac{g}{g+1}\left(Z'_{n+1}Z_{n+1}\right)^{-1} \qquad \tilde{\mu} = \tilde{V}Z'_{n+1}Y_n^*$$

where Z_{n+1} is the set of regressors that are included in model M_{n+1}.

4) Draw each of the components of Y_{n+1}^* from univariate truncated normal distributions as explained in Albert and Chib (1995).

We calculate the posterior probability of model M_j as the proportion of iterations that visit model M_j. Similarly, posterior means and credible intervals for θ or functions of θ (e.g. marginal effects) can be calculated using the draws obtained with the algorithm.

Coordination Clauses and the Price of Sovereign Debt

Ossip Robert Hühnerbein*

October 10, 2007

Abstract

Following recurring debt crises in the 1990s the Group of Ten has advocated the use of creditor coordination clauses as a means to enhance financial stability and making sovereign debt restructurings more orderly. While such clauses are common in bonds issued under English law, changes to the financial terms of a bond contract used to require unanimity under New York law. In recent years many emerging market borrowers have indeed included collective action clauses (CAC) in their sovereign issues. This paper analyzes the pricing of these clauses in the secondary markets. It shows that bonds without clauses to facilitate restructuring are viewed as being senior to CAC-bonds. In addition, it appears that the positive stability effects are increasing in the share of debt that is easy to coordinate. I interpret this as reflecting the fact that holders of non-coordinated bonds "freeride" on the obligingness of CAC bondholders.

Keywords: Sovereign Debt Crisis, Collective Action Clauses, Bondpricing

JEL classification: F33, F34, G12

*Munich Graduate School of Economics; Contact: ossip.huehnerbein@web.de. Financial support by the "Stability of the Global Financial System: Regulation and Policy Response", World Economy and Finance, ESRC Research Programme Phase I, Grant Ref. No.: RES-156-25-0026 is gratefully acknowledged.

1 Introduction

The financial crises of the 1990s have provoked a sizeable debate on what is referred to as the international financial architecture. After the resolution of the 1980s debt crisis with the Brady plan in the early 1990s it only took until 1994 when Mexico slit into the so called Tequila crisis. The surge of bond financing in emerging market economies led to financial crises of a new form. Creditor countries mostly borrowed at short maturities and had to roll over their claims frequently. If this coincided with a time of general market sentiment or if country fundamentals declined, the markets were unwilling to provide new money. Since the new money was needed to satisfy the outstanding debt the country experienced a roll-over crisis.

It was commonly understood that the problems arose because individual agents faced an incentive to *run for the exit* first, and thereby destabilized financial systems and aggravated matters. These first come first serve constraints were not only at work when sovereign debtors ran into liquidity based debt crises but also in balance of payments crises, currency crises, banking crises and combinations of these.[1] For the case of sovereign debt crises the discussion focused on the absence of any bankruptcy-like procedures for sovereign debtors and asked for measures to provide more orderly workouts for sovereigns in financial distress.[2]

The two main ideas discussed for reform of the international financial architecture were the contractual approach and the statutory approach. The contractual approach suggested that so called collective action clauses (CAC) should be embedded in sovereign bond contracts to overcome creditor coordination problems. These clauses stipulate that all proceeds a creditor receives from suing the debtor shall be split evenly among creditors (*sharing clause*), that only a prespecified threshold of creditors is needed to start litigation against the country (*non-acceleration clause*), and that the financial terms of the bond contract can be changed by a qualified majority of outstanding debt and that such an agreement is binding for dissenting creditors as well (*majority action clause*). While the latter solves hold out problems and facilitates restructuring the first two clauses are important to break the first come first serve constraint. The statutory approach consisted of the SDRM proposal by the IMF.[3] The SDRM would oversee sovereign restructurings similar to

[1]Overview on the East Asian Crisis in 1997 is found in Corsetti, Pesenti and Roubini (1999) and in Radelet and Sachs (1998), for self-fulfilling balance of payments and currency crisis see for example Obstfeld (1986,1996) and for banking and balance of payments *twin crises* see Kaminsky and Reinhard (1999).

[2]An initiating role can be attributed to Eichengreen and Portes (1995) and Sachs (1995).

[3]See Krueger (2002). The IMF also supported collective action clauses, see IMF (2002).

a domestic bankruptcy court. Both proposals drew heavy criticism. Coordination clauses were seen as undermining creditor rights and possibly putting the existing of the sovereign debt market at risk and the SDRM proposal was viewed as a rigid, bureaucratic international organization with doubtful legitimation.

Financial markets were wary towards the introduction of coordination clauses in sovereign bond contracts. On the creditor side there seemed not to be much consciousness on problems related to uncoordinated creditors. In contrast, a common opinion used to be that the status quo in sovereign debt restructuring, especially bond exchanges, allowed for sufficient renegotiation of stark claims.[4] Apparently fears of loosing bargaining power against emerging market borrowers preponderated discontent about problems with hold-out creditors or lengthy restructuring procedures. It is a common appraisal that the IMF's pressure towards a statutory mechanism strongly helped overcoming the resistance, as the private sector felt it had to opt quickly for the lesser of two evils.

In February 2003, Mexico was the first emerging market borrower to issue bonds including CAC under New York law. In the consecutive years many other countries have followed. This of course enables us to analyze whether sovereign debtors pay a premium for the option of facilitated restructuring. In the present analysis of secondary bond market spreads I disentangle this question into three effects:

- **Seniority Effect**
  Do investors require a premium for holding CAC bonds?
- **Coordination Effect**
  How does the fraction of bonds by one country that includes CAC affect the price of its outstanding bonds, i.e. how does the market perceive the effect of improved coordination of the outstanding debt?
- **Aggregation Effect**
  Does the pricing of CAC depend on whether only a small or a large share of debt is coordinated?

While the first question has been addressed frequently, I believe that the other two deserve special attention as well. The commonly cited positive and negative effects of collective action clauses are unlikely to depend on one single bond containing CACs or not. The positive effect, increased financial stability, assumes that a restructuring is faster and more equitable and efficient if creditors are coordinated. Furthermore it is less likely that pure panic driven or liquidity based crises occur. It seems

[4] See for example the remarks of Chamberlin (2002), executive director of the EMTA, at a conference of the Institute of International Finance, where he sums up views from the creditor side.

straightforward to assume that these effects depend on the amount of debt that contains the new clauses. The negative effect, moral hazard, implies that a country will tend to run reckless policies if the penalty in form of a crisis is less severe. Again, it seems natural to assume that the amount of coordinated debt determines the cost of the restructuring and hence the degree to which moral hazard is a concern. Both of these effects may very well apply for the holders of uncoordinated bonds as well. If fewer crises take place or if the economic losses associated with a crises are smaller due to more efficient restructuring mechanisms, holders of uncoordinated bonds will enjoy a more favourable payoff profile.[5] And in the case of increased moral hazard holders of bonds without CAC will face the consequence of unduly macro policies just as much as the CAC bondholders. Being aware of these interactions, the holders of sovereign bonds with CAC should be happy to see other creditors holding claims that are easy to restructure as well, because as mentioned above, the potential gains would otherwise partly accrue to creditors, who do not *pay* for such benefits by giving away some of their bargaining power.

The empirical analysis supports this view. I find evidence that bonds with collective action clauses are priced as if they were slightly junior to non CAC bonds. This penalty for holding instruments that are easy to restructure and therefore at higher risk of facing a writedown is more prominent for bonds with a higher probability of default, i.e. higher spreads. Evidence for the coordination effect is mixed. The results can however be explained that moral hazard reverses this effect for bad borrowers. The aggregation effect that CAC bondholders value the overall degree of coordination higher than other bondholders is very robust in the data.

The next section reviews the empirical literature on the pricing of collective action clauses. In section 3 I build a model to clarify my argumentation on the interaction of uncoordinated and coordinated bonds. The key argument is that CAC reduce the inefficiency arising from default, but worsen the negotiation position of CAC bondholders so that situations can arise were CAC bondholders face a larger haircut than other creditors do. I then address the question empirically. In section 4 and 5 I introduce the methodology and data. Section 6 contains the results and section 7 concludes.

[5]Holders of uncoordinated bonds could make their consent more expensive and thereby share some of the rents due to increased efficiency.

2 Overview of Existing Studies

Since the proposal on collective action clauses was tabled in the mid 1990s quite a few scholars have addressed the question on how the inclusion of such clauses into sovereign bond contracts would affect the market for sovereign debt. A fruitful coincidence for this task has been that bonds issued under English law traditionally include collective action clauses, while bonds issued under New York Law don't.

Eichengreen and Mody (2000) undertake an analysis of the spreads paid by emerging market issuers 1991-1998. Their analysis includes sovereign and corporate borrowers and they use UK governing law as a proxy for CAC inclusion. Since they examine launch spreads they have to control for endogeneity by running a first-stage regression explaining the choice of the governing law. Their main finding is that collective action clauses are priced differently depending on the quality of the borrower. When splitting their sample according to the credit rating they find that CACs reduce borrowing costs for more credit-worthy borrowers. Less credit-worthy creditors pay significantly higher spreads on their issues. They argue that while facilitated restructuring generates a benefit for all countries this is outweighed by moral hazard concerns for creditors with bad credit ratings.

Becker, Richards and Thaicharoen (2003) analyze primary market data as well as secondary market spreads. They find no evidence that CAC increase borrowing costs for creditors, independent of the creditworthiness of the borrower.[6] A lot of the results for the English law dummy are insignificant, no matter whether primary or secondary market spreads are analyzed. They see their results in line with anecdotal evidence that CAC inclusion seems to be no critical factor in financial circles when explaining variation of bond spreads. Accordingly, Becker, Richards and Thaicharoen (2003) conclude: "In summary, we consider it unlikely that governing law and the presence of CACs could have an impact [...] without market participants being acutely aware of this effect." (p. 26)[7]

Richards and Gugiatti (2003) examine primary as well as secondary market spreads. For the latter they find no significant effect of the inclusion of CACs. In their analysis on launch spreads they specifically analyze how a new issue was priced, not only given the inclusion of CAC, but also given the historical issue custom of

[6] Actually, Becker, Richards and Thaicharoen (2003) find the opposite effect to Eichengreen and Mody (2000) for some specifications: for comparatively bad borrowers CAC seem to have a positive effect, suggesting that the benefits of orderly restructuring are valued higher if this event is indeed more likely.

[7] Becker, Richards and Thaicharoen (2003) also question the magnitude of the effects Eichengreen and Mody (2000) claim. Their effects are in the order of 5-25 basis points while Eichengreen and Mody find effects of more than 100 basis points.

the country. They argue that if CACs were indeed valued negatively by the market, a change from issuing uncoordinated debt to issuing with CAC should increase returns and a cessation of the use of CAC should reduce the spread. Since they find no evidence for these effects they join Becker, Richards and Thaicharoen (2003) in claiming that coordination clauses are likely to be a dispensable factor in sovereign bond pricing. However, Richards and Gugiatti (2003) themselves question their event study on primary spreads. With 204 bonds issued between 1991 and 2001 by 10 countries it seems likely that switches between issues with and without CACs solely represent the conformance to market customs of borrowers seeking finance in London as well as New York respectively.[8]

The effect of aggregation, i.e. the composition of the outstanding bond debt has so far been addressed by Eichengreen and Mody (2003). However, they do not specifically focus on bonds with and without creditor coordination clauses. In an econometric assessment of launch spreads they analyze the effect of the number of bonds the debtor has already outstanding in the market. They find that more outstanding bonds result in marginally higher spreads. This supports the view that aggregation is priced in the market and that creditors are well aware of the fact that more issues will be more difficult to rearrange if necessary. Also, Eichengreen, Kletzer and Mody (2003) augment the dataset by Eichengreen and Mody (2000) by controlling for the amount of outstanding coordinated debt. Their results are insignificant, but when they split their sample according to creditor quality, they find that borrowers with unfavourable credit worthiness pay higher spreads on new CAC issues if they already have a lot of coordinated debt outstanding.

3 Theoretical Background

3.1 An Ex-Post Model of Debt Prices

The following model illustrates the interaction of bonds with and without coordination clauses in the secondary market for sovereign bonds. The model takes the composition of outstanding debt as given, i.e. the decision to issue bonds with or without collective action clauses is not modelled.

The total debt of the country outstanding is normalized to one. Of this a fraction γ is supposed to include collective action clauses. All bonds mature in the same period. The country enjoys full sovereign immunity, so that its repayment decision

[8] Further studies include Tsatsaronis (1999), Dixon and Wall (2000) and Petas and Rahman (1999). All these find that CAC do not raise borrowing costs significantly.

is voluntary. However, if the country does not honor its debt, creditors can impose a default pain on the country. If the country repays the creditors a fraction of the original debt and in turn the creditors lift the default pain, I call this arrangement a debt restructuring, which is a Pareto improvement. I do not analyze the negotiation process that leads up to this agreement but claim that by the threat of the default claim a fraction (θ) of output (y) is *pledgeable*.[9] Formally the country will repay up to θy.

The negotiation of a sovereign debt restructuring is costly in the model. Due to the prolonged state of uncertainty until an agreement is reached during which the country suffers the consequences of default, a fraction of output is lost. The literature usually motivates this with difficulties in international trade and domestic economic stress arising from a sovereign default. However, I assume that for bonds with collective action clauses these effects do not apply. If the country sticks to the rules defined in the bond contract, the country never defaults on outstanding claims. For example the exercise of a majority vote upon a rescheduling, does not allow creditors to sue the country.[10] In the model the use of collective action clauses for restructuring is costless, while restructuring claims without CAC bears a loss of output α. The cost of restructuring is of course proportional to the amount of uncoordinated debt outstanding. So disposable income is:

$$\tilde{y} = (1 - (1 - \gamma)\alpha)y \tag{1}$$

If the country arrives in the repayment period it has to decide whether to repay or to seek a restructuring. The country can also decide to restructure only one type of claim.[11] The returns from either decision can be summarized in the following table 1:

To make a useful comparison of the payoffs we have to clarify how high the restructured repayment R_i is. Therefore we use the fact that the pledgeable output is $\theta(1-(1-\gamma)\alpha)y$. It follows that $R_3 = \theta(1-(1-\gamma)\alpha)y$. How much will the country offer in a restructuring to a fraction of bondholders? Just the same amount. While a higher amount makes no sense from the debtors point of view, a lower offer would be rejected by the creditors, because they know they can get more by rejecting

[9] See Fernandez and Rosenthal (1990) on the negotiation process with limited enforcement.

[10] The procedure of suing is fruitless for the creditor, but essential in making the default costly.

[11] This is a critical assumption since we do not perceive selective defaults regularly in the sovereign bond market. Nevertheless, Russia and Argentina attempted to treat issues held mainly by domestic residents more favourably. An alternative view supporting the payoff structure in the model is, that holders of uncoordinated bonds have the option of selling to holdout creditors. So the excess payoff uncoordinated creditors gain can be described as a holdout premium.

Decision	Payoff
repay all	$y - 1$
restructure just CAC bonds	$y - \gamma R_1 - (1-\gamma)$
restructure just no-CAC bonds	$(1 - \alpha(1-\gamma))y - \gamma - (1-\gamma)R_2$
restructure all	$(1 - \alpha(1-\gamma))y - R_3$

Table 1: Country Payoffs

and forcing the country into a full restructuring. With $R_1 = R_2 = R_3 = R$ it is straightforward to see that the country will never seek a restructuring just with the holders of uncoordinated bonds. The loss of output is the same if all debt is reduced. However, seeking agreement on a debt reduction only with the holders of CAC bonds makes sense. The fact that restructuring bears no cost makes these creditors easy prey of a country unwilling to repay. If they are offered a reduced claim R, they have no means to ask for more. Forcing the country into default does not promise any excess payments. The resulting payoffs for the country are summarized in table 2. For simplicity it is convenient to write the effect of a restructuring on output as $\beta = 1 - (1-\gamma)\alpha$.

Decision	Payoff
repay all *(repay)*	$y - 1$
restructure just CAC bonds *(cac)*	$(1 - \gamma\theta\beta)y - (1-\gamma)$
restructure all *(default)*	$(1-\theta)\beta y$

Table 2: Country Payoffs

Let output be distributed randomly with a support of $[0, \infty]$ and some general density $f(y)$. Then we can derive decisions for the country depending on the realization of y. First, it is easy to see that the country will restructure all debt if $y \to 0$. Also, we can derive that for $y \to \infty$ the country will repay the claims of both types of creditors. To see that there is an intermediate zone in which the country only restructures the bonds that include coordination clauses, first note that the slope of the equation describing the payoff from restructuring only the CAC bonds is inbetween the other two (*repay,default*). Firstly, $1 - \gamma\theta\beta < 1$, so that repaying all bonds becomes more attractive than restructuring just the CAC bonds for some y. Also, it can be shown that a separate restructuring of CAC bonds is better than restructuring all debt for some y:

$$1 - \gamma\theta\beta > (1-\theta)\beta \tag{2}$$

$$\Leftrightarrow \theta\beta(1-\gamma) > 0 \tag{3}$$

In order to prove the existence of a range for y where restructuring just the CAC bonds is optimal, we have to prove that the intersection of the $default$ and the cac line is for smaller values of y than the intersection between the cac and the $repay$ line. Let us define the intersections:

$$y_1 := (1-\theta)\beta y_1 = (1-\gamma\theta\beta)y_1 - (1-\gamma) \quad (4)$$

$$y_2 := (1-\gamma\theta\beta)y_1 - (1-\gamma) = y_1 - 1 \quad (5)$$

We solve for these intersections and see that there exists a range for y such that restructuring only CAC claims is optimal.

$$y_1 = \frac{1}{\alpha + \beta\theta} < \frac{1}{\theta\beta} = y_2 \quad (6)$$

Since α is posititive $y_2 > y_1$ holds unambiguosly. These findings are illustrated in figure 3.1. Let us know analyze how the different types of bonds will be priced in the

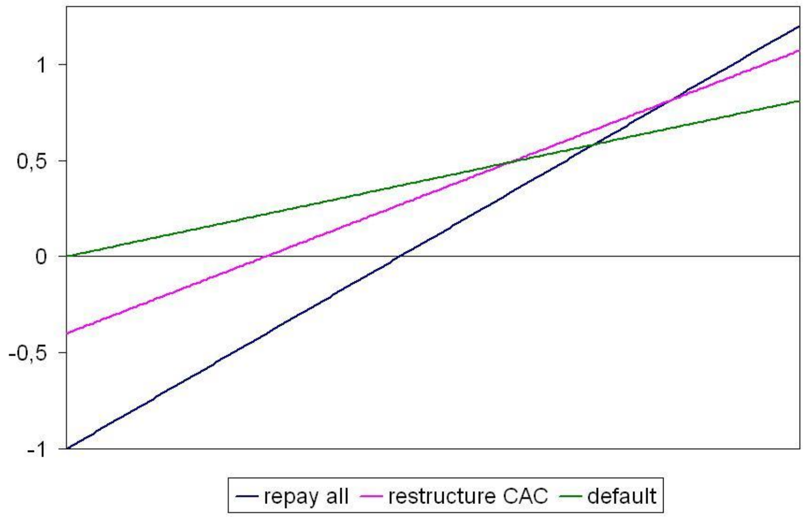


Figure 1: Debtor Payoffs

secondary market. I assume that the benchmark interest rate is normalized to zero. So a risk-free bond in the model, that promises a return of one dollar, will be priced exactly at one dollar. The interest rate of the sovereign bonds is the fraction of the contracted repayment over the current price minus one. Since the discount factor is one, the current price equals the expected return. And since the risk-free interest rate is zero, the interest rate and the spread are analogous. With s_{CAC} I denote the interest rate differential of bonds including CAC to uncoordinated bonds.

$$i_{CAC} = \frac{1}{P_{CAC}} - 1 \tag{7}$$

$$i_{NO} = \frac{1}{P_{NO}} - 1 \tag{8}$$

$$s_{CAC} = i_{CAC} - i_{NO} = \frac{1}{P_{CAC}} - \frac{1}{P_{NO}} \tag{9}$$

The price of each bond is just its expected return. These are given by:

$$P_{CAC} = \int_0^{y_2} \beta\theta y f(y)dy + \int_{y_2}^{\infty} f(y)dy \tag{10}$$

$$P_{NO} = \int_0^{y_1} \beta\theta y f(y)dy + \int_{y_1}^{\infty} f(y)dy \tag{11}$$

Since $\beta\theta y < 1$ and $y_1 < y_2$ we can immediately propose that:

Proposition 1. *The spreads of the sovereign bonds over a risk-free asset are strictly positive. Collective action clause bonds are priced with a higher spread than uncoordinated bonds.*

$$i_{CAC} > 0,\ i_{NO} > 0 \tag{12}$$

$$s_{CAC} > 0 \tag{13}$$

Our next task is to analyze how a variation in the amount of coordinated debt affects the spreads of the uncoordinated as well as the coordinated sovereign bonds.

$$\frac{\partial i_{CAC}}{\partial \gamma} = -\left(\frac{1}{P_{CAC}}\right)^2 \frac{\partial P_{CAC}}{\partial \gamma} \tag{14}$$

$$\frac{\partial i_{NO}}{\partial \gamma} = -\left(\frac{1}{P_{NO}}\right)^2 \frac{\partial P_{NO}}{\partial \gamma} \tag{15}$$

The partial derivatives of the prices are:

$$\begin{aligned}
\frac{\partial P_{CAC}}{\partial \gamma} &= \frac{\partial y_2}{\partial \gamma}\beta\theta y_2 f(y_2) + \int_0^{y_2} \frac{\partial \beta}{\partial \gamma}\theta y f(y) dy - \frac{\partial y_2}{\partial \gamma} f(y_2) \\
&= -\frac{\partial y_2}{\partial \gamma} f(y_2)(1-\beta\theta y_2) + \int_0^{y_2} \frac{\partial \beta}{\partial \gamma}\theta y f(y) dy \\
&= \int_0^{y_2} \frac{\partial \beta}{\partial \gamma}\theta y f(y) dy \qquad (16) \\
\frac{\partial P_{NO}}{\partial \gamma} &= \frac{\partial y_1}{\partial \gamma}\beta\theta y_1 f(y_1) + \int_0^{y_1} \frac{\partial \beta}{\partial \gamma}\theta y f(y) dy - \frac{\partial y_1}{\partial \gamma} f(y_1) \\
&= -\frac{\partial y_1}{\partial \gamma} f(y_1)(1-\beta\theta y_1) + \int_0^{y_1} \frac{\partial \beta}{\partial \gamma}\theta y f(y) dy \qquad (17)
\end{aligned}$$

For the analysis of the partial derivatives of the intersections with respect to γ we have:

$$\delta_1 = \frac{\partial y_1}{\partial \gamma} = -\frac{\alpha\theta}{(\alpha + (1-\alpha(1-\gamma))\theta)^2} = -\frac{\alpha\theta}{(\alpha+\beta\theta)^2} \qquad (18)$$

$$\delta_2 = \frac{\partial y_2}{\partial \gamma} = -\frac{\alpha}{\theta(1-\alpha(1-\gamma))^2} = -\frac{\alpha}{(\beta\theta)^2} \qquad (19)$$

Since α and θ are within the interval $(0, 1)$, the partial derivatives of the intersections with respect to the fraction of CAC bonds (δ_1, δ_2) are negative and we can state another proposition.

Proposition 2. *The spread a holder of a sovereign bond demands in relation to a risk-free bond is decreasing in the fraction of sovereign debt endowed with collective action clauses. This effect arises for both the uncoordinated bonds and the bonds including collective action clauses themselves.*

$$\frac{\partial i_{CAC}}{\partial \gamma} < 0 \qquad (20)$$

$$\frac{\partial i_{NO}}{\partial \gamma} < 0 \qquad (21)$$

Furthermore, we might be interested in knowing whether the two types of bonds are affected differently by a variation in the amount of debt that can be restructured easily. Therefore we have to compute the partial derivative of s_{CAC} with respect to γ.

$$\frac{\partial s_{CAC}}{\partial \gamma} = -\left(\frac{1}{P_{CAC}}\right)^2 \frac{\partial P_{CAC}}{\partial \gamma} + \left(\frac{1}{P_{NO}}\right)^2 \frac{\partial P_{NO}}{\partial \gamma} \qquad (22)$$

If we make a general assumption on the distribution of y we can unambiguously state:

Proposition 3. *If $f(y)$ describes a single humped distribution and the maximum density is for $y > y_2$, the spread between sovereign bonds by the same issuer with and without collective action clauses is declining when the fraction of bonds with coordination clauses rises.*

$$\frac{\partial s_{CAC}}{\partial \gamma} < 0 \tag{23}$$

For symmetric distributions of y the assumption made implies that the probability of a restructuring must be below 50%. Depending on the expected loss in the event of a restructuring this requires that spreads must be very high to fall short of this assumption. For example for an expected haircut of 50% the assumption is valid for spreads below 2500 basis points.

Proof. If we make use of the fact that $P_{NO} > P_{CAC}$ we have to show that:

$$\frac{\partial P_{CAC}}{\partial \gamma} > \frac{\partial P_{NO}}{\partial \gamma} \tag{24}$$

Substituting in yields:

$$\begin{aligned} \int_{y_1}^{y_2} \frac{\partial \beta}{\partial \gamma} \theta y f(y) dy &> -\frac{\partial y_1}{\partial \gamma} f(y_1)(1 - \beta\theta y_1) \\ \int_{y_1}^{y_2} \alpha\theta y f(y) dy &> \frac{\alpha\theta}{(\alpha+\beta\theta)^2} f(y_1)(1 - \frac{\beta\theta}{\alpha+\beta\theta}) \\ \int_{y_1}^{y_2} y f(y) dy &> \frac{1}{(\alpha+\beta\theta)^2} f(y_1) \frac{\alpha}{\alpha+\beta\theta} \end{aligned} \tag{25}$$

Since $f(y)$ is single-humped with its maximum for $y > y_2$ we know that $f(y_1) \leq f(y)$ for all $y \in [y_1, y_2]$. Therefore we can conclude that inequality 25 will always hold if:

$$f(y_1) \int_{y_1}^{y_2} y dy > \frac{\alpha}{(\alpha+\beta\theta)^3} f(y_1) \tag{26}$$

Solving for the integral we get:

$$\begin{aligned} \frac{1}{2}\left[\frac{\alpha^2 + 2\alpha\beta\theta}{\beta^2\theta^2(\alpha+\beta\theta)^2}\right] &> \frac{\alpha}{(\alpha+\beta\theta)^3} \\ \frac{1}{2}\left[\frac{\alpha + 2\beta\theta}{\beta^2\theta^2}\right] &> \frac{1}{\alpha+\beta\theta} \end{aligned} \tag{27}$$

Since α is positive inequality 27 will surely hold if:

$$\frac{1}{2}\left[\frac{\alpha + 2\beta\theta}{\beta^2\theta^2}\right] > \frac{1}{\beta\theta}$$
$$\frac{1}{2}\left[\frac{\alpha + 2\beta\theta}{\beta\theta}\right] > 1$$
$$1 + \frac{\alpha}{2\beta\theta} > 1$$

□

Next we address the question whether the effects identified so far vary systematically with the overall quality of the debtor. A meaningful way to distinguish between good and bad debtors in the model is the distribution of y. I assume that good countries' output is distributed according to the density $f(y)$, whereas bad debtors' output is distributed with a density $\tilde{f}(y)$. I assume that $\tilde{F}(y_2) > F(y_2)$ which implies that it is more likely that a bad country refuses to honor its debt. Furthermore, I assume that $\tilde{f}(y) > f(y)$ for all $y \in (0, y_2)$ so that every single state where the country dishonors the debt is more likely for bad debtors.[12]

Proposition 4. *The premium creditors pay for uncoordinated bonds is higher for comparatively bad debtors.*

$$s_{CAC}|\tilde{f}(y) > s_{CAC}|f(y) \tag{28}$$

Proof. From equation 9 we know that the spread is proportional to the difference between the prices of the two bonds. From equations 10 and 11 the difference between bondprices is:

$$\int_{y_1}^{y_2} (1 - \beta\theta y) f(y) dy \tag{29}$$

Since $\tilde{f}(y) > f(y)$ for $y \in [y_1, y_2]$ the difference between prices is larger for worse debtors. Henceforth, the spread s_{CAC} of bonds with CAC over other bonds by the same issuer is also larger. □

Proposition 5. *The sensitivity of the spread of a sovereign bond over a riskless bond with respect to the degree of coordination γ is higher for comparatively bad*

[12]For single-humped distributions this includes a squeezing of the density towards smaller values as long as the maximum density is for $y > y_2$.

debtors.

$$\left.\frac{\partial i_j}{\partial \gamma}\right| \tilde{f}(y) > \left.\frac{\partial i_j}{\partial \gamma}\right| f(y) \ , \qquad j \in (CAC, NO) \tag{30}$$

Proof. Since the prices are smaller for worse debtors we have to proof that the derivatives are larger, i.e.:

$$\left.\frac{\partial P_j}{\partial \gamma}\right| \tilde{f}(y) > \left.\frac{\partial P_j}{\partial \gamma}\right| f(y) \ , \qquad j \in (CAC, NO) \tag{31}$$

We have from equations 16 and 17:

$$\frac{\partial P_{CAC}}{\partial \gamma} = \int_0^{y_2} \alpha\theta y f(y) dy$$
$$\frac{\partial P_{NO}}{\partial \gamma} = -\delta_1 f(y_1)(1 - \beta\theta y_1) + \int_0^{y_1} \alpha\theta y f(y) dy$$

Since $\tilde{f}(y) > f(y)$ for $y < y_2$ the derivatives of the prices with respect to the degree of coordination are larger for bad debtors. □

To set out for a proposition on the behaviour of proposition 3 for bad debtors we have to make an extra assumption:

$$g(y) := \tilde{f}(y) - f(y)$$
$$g(y_1) \leq g(y) \ , \quad \text{for } y \in (y_1, y_2) \tag{32}$$

To grasp the intuition for this assumption recall from equations 16 and 17 how a change of γ affects the prices of the two types of bonds. For the CAC bonds a change in γ raises β so that recovery is better for all default states $y < y_2$. The marginal effect cancels out, because $\theta\beta y_2 = 1$. For bonds without CAC the effect of γ on default states is smaller simply because there are less default states. However, the marginal change of y_1 has a positive effect because the return jumps from $\frac{\beta\theta}{\alpha+\beta\theta}$ to 1. For this effect at y_1 not to outweigh the accumulated effects between y_1 and y_2 we needed that the density at y_1 is smaller than for $y \in (y_1, y_2)$. A similar assumption about the change of the density has to be made for proposition 6.[13]

[13]Two things can be said on the restrictiveness of this assumption. Firstly, if the density of a normal distribution is shifted to the left, assumption 32 holds for default probabilities below 50% and well beyond. In so far the assumption includes what one could have in mind as a benchmark example. Secondly, imagine the situation in real world financial markets. With some uncertainty about the parameter values, the point density $f(y_1)$ seems to loose weight towards the area $\int_{y_1}^{y_2} f(y)dy$. For the marginal effect that the assumption in equation 32 disarms to be effective, creditors had to know that exactly the density at the switching point from defaulting on both types of claims to restructuring only CAC bonds, would increase

Proposition 6. *The sensitivity of the spread of a sovereign bond with CAC over a sovereign bond without CAC with respect to the degree of coordination γ is higher for comparatively bad debtors.*

$$s_{CAC}|\tilde{f}(y) < s_{CAC}|f(y) \tag{33}$$

Proof. Recall the definition of $\frac{\partial s_{CAC}}{\partial \gamma}$ from equation 22:

$$\frac{\partial s_{CAC}}{\partial \gamma} = -\left(\frac{1}{P_{CAC}}\right)^2 \frac{\partial P_{CAC}}{\partial \gamma} + \left(\frac{1}{P_{NO}}\right)^2 \frac{\partial P_{NO}}{\partial \gamma}$$

From Proposition 4 we know that the difference between prices increases when the countries' prospect worsens. So we are left to show:

$$\left(\frac{\partial P_{CAC}}{\partial \gamma} - \frac{\partial P_{NO}}{\partial \gamma}\right)\Bigg| \tilde{f}(y) > \left(\frac{\partial P_{CAC}}{\partial \gamma} - \frac{\partial P_{NO}}{\partial \gamma}\right)\Bigg| f(y) \tag{34}$$

From equation 25 we can write $\frac{\partial P_{CAC}}{\partial \gamma} - \frac{\partial P_{NO}}{\partial \gamma}$ as:

$$\int_{y_1}^{y_2} \alpha\theta y g(y) dy + \frac{\partial y_1}{\partial \gamma} g(y_1)(1 - \beta\theta y_1) > 0$$

This is the same equation as in 25 with $g(y)$ instead of $f(y)$. With the assumption made on $g(y)$ the proof of proposition 3 holds here as well. □

The model just presented delivers three testable predictions about the pricing of sovereign bonds with and without collective action:

- Seniority Effect: Bonds that include collective action clauses have a positive spread over equivalent bonds without these clauses. This effect arises because facilitated restructuring will make a writedown in debt more likely for these bonds.
- Coordination Effect: Sovereign bond spreads are lower if the fraction of coordinated debt is larger. This stems from the fact that restructurings are more orderly if more debt includes collective action clauses.
- Aggregation Effect: The spread between bonds with and without coordination clauses is decreasing in the fraction of outstanding debt that is coordinated. Since the likelihood of a restructuring is higher for the bonds with facilitated

more than the probability that only CAC bonds be restructured. If one assumes that real financial markets act with some blurriness this is a seemingly far fetched idea.

restructuring, the holders of these bonds benefit comparatively more from more efficient workouts.

Furthermore all three effects are shown to be larger in magnitude for debtor countries with a comparatively worse distribution of pledgeable output.

3.2 Moral Hazard

The model presented in the previous subsection takes the distribution of output as given. However, the theory on the borrowing of sovereign debtors puts a lot of emphasis on the policy decision of the debtor and the associated incentives. This moral hazard concern highlights that the debt contract has to be disciplining enough to secure repayment to the creditors. In the context of creditor coordination clauses it has been argued that facilitating restructuring by improving creditor coordination will increase the likelihood of default. This could eventually lead to a breakdown of the market for sovereign debt because creditors would be unwilling to provide funds given the high probability of default.

So how would the model sketched above be affected by moral hazard? First, it is important to note, that moral hazard is a countrywide phenomenon. In so far as the inclusion of CAC may affect the repayment decision for each individual bond issue, it is captured in the model above. I assume that the distribution of pledgeable income depends on the expected cost of a restructuring for the country. This moral hazard affects the holders of CAC bonds and other bonds of the same creditor the same way.

Formally, let us state that the distribution of output $f(y)$ depends on γ such that:

$$\frac{\partial f(y)}{\partial \gamma} = g(y|m) \text{ , with} \tag{35}$$

$$g(y|m) \begin{cases} < 0 \text{ , if } y < y_2 \\ > 0 \text{ , if } y \geq y_2 \end{cases} \tag{36}$$

where m is a parameter for the degree of moral hazard such that

$$\frac{\partial g(y|m)}{\partial m} \begin{cases} < 0 \text{ , if } y < y_2 \\ > 0 \text{ , if } y \geq y_2 \end{cases} \tag{37}$$

Given these structure we can rewrite the derivatives of the bond price with respect to the degree of coordination from equations 16 and 17:

$$\frac{\partial P_{CAC}}{\partial \gamma} = \int_0^{y_2} \alpha\theta y f(y) dy + \underbrace{\int_0^{y_2} \beta\theta y g(y|m) dy + \int_{y_2}^{\infty} g(y|m) dy}_{\text{Moral Hazard}} \tag{38}$$

$$\frac{\partial P_{NO}}{\partial \gamma} = \int_0^{y_1} \alpha\theta y f(y) dy - \delta_1 f(y_1)(1 - \beta\theta y_1) + \underbrace{\int_0^{y_1} \beta\theta y g(y|m) dy + \int_{y_1}^{\infty} g(y|m) dy}_{\text{Moral Hazard}} \tag{39}$$

Since $G(y|m) = 0$ we can rewrite the moral hazard component as:

$$\int_0^{y_j} (\beta\theta y - 1) g(y|m) dy < 0 \tag{40}$$

where we use the fact that any additional density for $y < y_2$ must correspond to a loss of density for $y > y_2$. So if $g(y|m)$ is large enough, an increase in the degree of coordination may decrease prices and increase spreads. It follows that a sufficient amount of moral hazard can reverse the results from proposition 2 and 5.

4 Methodology

The standard regression to estimate the determinants of bond spreads is of the form:[14]

$$log(spread) = B\mathbf{X} + \epsilon \tag{41}$$

The dependent variable is the log of the spread. The spread is difference between the interest rate curve implied by the current bond price and the future contracted repayments and a benchmark yield curve. This is explained by a set of characteristics **X**. **X** contains information about the country, for example debt ratios, but also about the issue, e.g. the duration and the currency denomination. Lastly, some characteristics are not specific to the bond, especially I include time dummies. The disturbance term ϵ is a random error. General country characteristics in my regression equation include GDP growth, inflation, GDP per capita and current account balance. Explaining variables addressing the external indebtedness of the

[14]See Eichengreen an Mody (2000)

country are the debt/GDP ratio, debt/reserves, debt/exports and a ratio of short term to long-term debt. I also include a dummy for the currency in which the bond in denominated. As mentioned above I have seven points in time and use dummies to account for a general movement in the sovereign bond market. To account for the duration of a bond I include the modified duration both linear and quadratic. To assess the various predictions on the inclusion of collective action clauses I include a dummy CAC_i, indicating whether the bond entails coordination clauses. Furthermore I construct a variable to reflect the aggregated coordination facility for a country at any time. With VOL_{it} being the volume of bond i active at time t I define:

$$FRAC_t = \frac{\sum_i CAC_i VOL_{it}}{\sum_i VOL_{it}} \tag{42}$$

To account for differences in the pricing of aggregation for bonds that include CACs and bonds which don't I also include an interaction variable $CAC * FRAC$.

$$log(S) = \beta_0 + \beta_1 CAC + \beta_2 FRAC + \beta_3 CAC * FRAC + B\mathbf{X} + \epsilon \tag{43}$$

In contrast to previous studies I do not include country ratings such as those provided by Standard&Poors or Moody's. Such an assessment would inevitably regard the structure of the debt in terms of long-term versus short-term debt and consider the effects of facilitated restructuring through collective action clauses. Since the ranking is countrywide this is no problem in analyzing the pricing of the clauses (seniority effect), but will affect the pricing of the debt composition (coordination and aggregation effects).

I estimate the model by ordinary least squares. Eichengreen and Mody (2000) note that this may generate biased results if the choice on the inclusion of coordination clauses is endogenous. Two things assure me that endogeneity is not a problem in the regressions I run: first, as Becker, Richards and Thaicharoen (2003) argue, there are fewer problems with endogeneity in an analysis of secondary market spreads. At the point in time the investigation is exercised the decision on contractual features is long in the past so that it can be viewed as exogenous. Second, in my sample all countries have issued bonds with collective action clauses even under New York law. Therefore, country characteristics could hardly serve an explanation for endogeneity.[15] This also strengthens the first argument, because at most points

[15]Richards and Guigiatti (2003) restrict their sample to countries that have issued with and without CAC to check whether their results are robust to endogeneity problems.

in time I compare bonds with and without collective action clauses from the same issuers.

5 Data

The bond data is obtained from DATASTREAM. The dataset contains of 436 bonds issued by 19 countries that mature after April 2003. These include Belize, Brazil, Chile, Colombia, Costa Rica, Guatemala, Indonesia, South Korea, Mexico, Panama, Peru, the Philippines, Poland, South Africa, Turkey, Uruguay, and Venezuela. The CAC variable is generated following the issue date, with February 2003 marking the change. I use the information on the 19 countries reported in Drage and Hovaguimian (2004). They state:

> Thus, it now appears that [...] the inclusion of CACs in sovereign bonds issued under New York law has switched from being the exception to becoming the market standard.

Since the countries in my sample have all issued New York law bonds including CAC, I suspect that they did not return to issuing bonds without coordination clauses for later issues, where I don't have this information. Nevertheless, I think that issuance after February 2003 for New York law bonds serves as a very good instrument for CAC inclusion. For bonds issued under English law I follow Eichengreen and Mody (2000) and presume that they always include CAC. The sample also includes some bonds issued under German, Luxembourg, Swiss and Japanese law respectively. The treatment of these bonds is less obvious. In my main specification I suspect these bonds to include CAC after February 2003 just like the New York law bonds, but I also test the predictions if I assume the opposite or exclude these bonds from the sample. Macroeconomic Data is taken from the International Monetary Fund International Finance Statistics, the Joint External Debt Hub jointly developed by the Bank for International Settlements (BIS), the International Monetary Fund (IMF), the Organization for Economic Cooperation and Development (OECD) and the World Bank (WB) and the IMF World Economic Outlook (WEO). The first two sources are quarterly and the BIS is annually. In order not to stress the data too much but still allow for enough variation of the debt composition for each country, I construct a semi-annual dataset from 2003-1 to 2006-1. Figure 2 depicts the development of the share of bonds with CAC in my sample.

The total number of observations is 1862 if you drop all bonds that have no data at that date. The reduction to 1281 observations is due to some missing data in

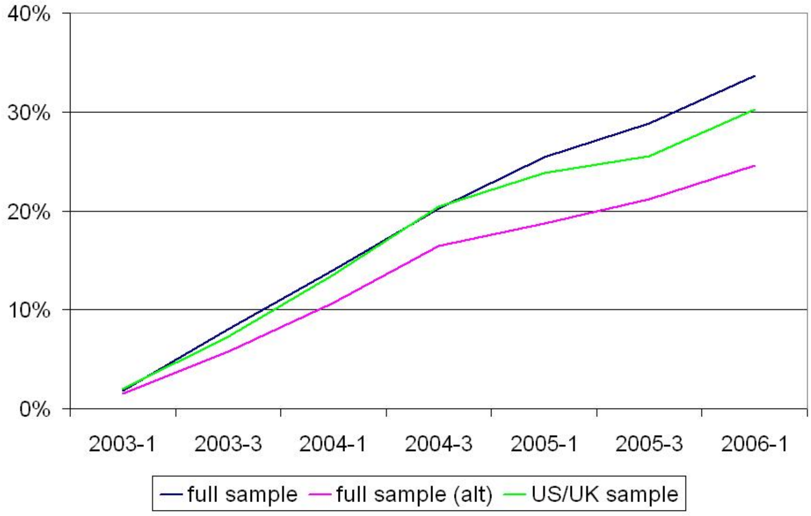


Figure 2: Share of Bonds with CAC

the IFS as well as the JEDH data. In table 3 I report descriptive statistics for some variables. There are only 13 observations with spreads over 25% in the sample of which 11 belong to Uruguay. Deleting these outliers does not substantially alter results.

Variable	Mean	Minimum	Maximum
Spread (%)	3.86	0.015	84.69
Duration (years)	5.30	0.008	20.65
FRAC	0.23	0	0.96
GDP growth (%)	4.41	-11	18.9
GDP per capita (US$)	8.185	3.697	20.590
Inflation (%)	8.95	0.2	44.8
Current Account (%)	-0.13	-22.3	19.1
Debt/GDP	0.1575	0.0095	0.6006
Debt/Reserves	2.234	0.060	6.131
Debt/Exports	3.237	0.125	30.58
Short-Term/Long-Term	0.095	0	0.802

Table 3: Descriptive Statistics

6 Results

6.1 General

The first regression includes all active bonds, e.g. all bonds that were not yet repaid at the respective date. The duration has a highly significant positive effect on the spread. This reflects an upward sloping yield curve that is expected for sovereign bonds. Poorer countries pay higher interest rates as the significantly negative coefficient on GDP per capita suggests. All debt ratios have positive coefficients as expected. However, the classical debt/GDP ratio is insignificant at the 5%-level. Inflation has a significant positive effect on the spread. This comes with no surprise as inflation often goes with real depreciation that devaluates domestic income. The significant influence of the ratio of short to long term debt shows that markets are well aware of the risk excessive maturity mismatch imposes. The current account balance and GDP growth have the expected negative sign but are insignificant. I also include two variables to control for the liquidity. The volume of an issue indeed seems to reduce the spread of the bond. This reflects the more favourable pricing of more liquid bonds. The variable *on the run* captures whether a bond has been launched within the last six month. Often, the most recently issued bond with certain characteristics is the most liquid one. It is especially important to control for this issue in this context, since most bonds bearing CAC's are introduced within the analyzed period.[16]

The variables of special interest are those associated with the coordination features of the outstanding debt. The seniority effect would imply that the variable CAC has a positive impact on the spread. This is indeed the case. The coordination effect would suggest that all creditors of a country, no matter whether their issues include CAC, enjoy a large share of coordinated debt. Consequently, the coefficient of FRAC should be negative. This result is present and significant. Lastly, the aggregation effect suggests that the holders of CAC bonds are willing to pay a premium for other issues being coordinated as well. The implied negative coefficient of CAC$*$FRAC shows up significantly.

As a robustness check I present results from alternative regressions with country dummies in the appendix. The dataset only spans four consecutive years. This may limit the variation of macro data on the country level over time. It may therefore be useful to substitute the set of explanatory macroeconomic variables

[16]See for example Warga (1992), Jankowitsch et al (2006), Fleming (2003) and Elton and Green (1998) on liquidity effects in the bond market.

explanatory variable	Coefficient	t-Statistic	
Duration	0.2800	14.32	***
Duration^2	-0.0145	-9.90	***
CAC	0.1971	2.26	**
FRAC	-0.6645	-2.31	**
CAC∗FRAC	-1.2571	-5.03	***
Euro	-0.0713	-1.62	
Yen	0.3223	3.12	***
£	-0.0941	-0.79	
SF	0.6581	3.77	***
2003 Q1	-0.0566	-0.47	
2003 Q3	-0.1764	-2.15	**
2004 Q1	-0.3676	-4.75	***
2005 Q1	-0.0935	-1.47	
2005 Q3	-0.2749	-3.81	***
2006 Q1	-0.3644	-4.30	***
GDP growth	-0.0153	-1.10	
Current Account	-0.0185	-1.50	
Inflation	0.0302	6.81	***
log(GDP per capita)	-0.8191	-4.87	***
Debt/GDP	1.6419	1.59	
Debt/Exports	0.1924	3.77	***
Debt/Reserves	0.1378	2.94	***
Short-Term/Long-Term	1.7465	4.04	***
log(Volume)	-0.0767	-3.36	***
on the run (6 month)	-0.0999	-1.23	
Constant	11.6758	7.44	***
Adjusted R^2	0.5885		
Observations:	1282		

Table 4: Regression 1

Dependent variable is log(spread). *,**,*** denote significance at the 10%,5% and 1% level respectively.

with country dummies. In the appendix I present the results from this exercise and from regressions where I simply add the country dummies to the regressions I present in this section. The results are in table 6.

6.2 Structural Breaks

In this subsection I analyze whether the effects of collective action clauses on the pricing of bonds in the secondary market varies in different circumstances. I concentrate on the question whether CAC are priced differently in stressful times. Therefore I construct a dummy to denote that a country experiences bad times in financial markets, BAD. I then run the regression equation as before but include interaction terms for the three coordination variables with the dummy BAD.

$$\begin{aligned} log(S) = &\beta_0 + \beta_1 CAC + \beta_2 FRAC + \beta_3 CAC * FRAC + \\ &\beta_4 BAD * CAC + \beta_5 BAD * FRAC + \beta_6 BAD * CAC * FRAC + B\mathbf{X} + \epsilon \end{aligned}$$

The coefficients β_4,β_5 and β_6 report differences in the effect of CAC,$FRAC$ and $CAC{*}FRAC$ in the bad sample, i.e. how the seniority, coordination and aggregation effects occur in the sample with less creditworthy creditors. The question is of course, how to construct the sample of *bad* countries. Since the spread is a measure of the default probability it is natural to split the sample according to the level of the spread. However, the spread is not only influenced by characteristics of the country but also by issue characteristics. To control for these I first construct a fit from the first regression, where I include Duration, Duration^2, the currency dummies and liquidity terms. I then construct the spread over fit to get the spread that is not explained by issue characteristics. The set of bad countries is then defined as those observations, where the spread is much higher than expected from issue characteristics. This procedure prevents me from defining most of the long-term bonds as BAD due to their high spread. The results from this regression are reported in table 5.

The sample of countries with higher spreads is one where the probability of default, or more precisely the expected loss due to default, is higher compared to the rest of observations. This is the situation described in section 3.1 where the distribution of pledgeable output is less favourable. Consequently, propositions 4 to 6 should guide us as to what results we would expect. The three effect should all be more pronounced than in the sample of countries with better repayment prospects.

For the seniority effect, measured by the coefficient on $BAD * CAC$ I find no

explanatory variable	Coefficient	t-Statistic	
Duration	0.2176	13.00	***
Duration^2	-0.0109	-8.71	***
CAC	0.1355	1.43	
FRAC	-1.2032	-4.64	***
CAC∗FRAC	-0.0571	-0.20	
BAD∗CAC	-0.0865	-0.67	
BAD∗FRAC	2.0334	6.61	***
BAD∗CAC∗FRAC	-1.5156	-3.70	***
Euro	-0.1442	-3.88	***
Yen	0.0664	0.76	
£	-0.2127	-2.12	**
SF	0.1445	0.97	
2003 Q1	0.1825	1.78	*
2003 Q3	0.0021	0.03	
2004 Q1	-0.1428	-2.18	**
2005 Q1	0.0181	0.34	
2005 Q3	-0.0723	-1.17	
2006 Q1	-0.1133	-1.55	
GDP growth	-0.0111	-0.95	
Current Account	-0.0345	-3.34	***
Inflation	0.0142	3.73	***
log(GDP per capita)	-0.5965	-4.23	***
Debt/GDP	1.0276	1.05	
Debt/Exports	0.1140	2.42	**
Debt/Reserves	0.0183	0.45	
Short-Term/Long-Term	0.8053	2.21	**
log(Volume)	-0.0400	-2.08	**
on the run (6 month)	-0.0906	-1.33	
constant (BAD)	0.6800	9.32	***
Constant	9.7144	7.37	***
Adjusted R^2	0.7136		
Observations	1282		

Table 5: Effects for countries with high spreads
Dependent variable is log(spread). *,** and *** denote significance at the 10%,5% and 1% level respectively.

significant clues in the data. For the aggregation effect (proposition 3 and 6) we also find evidence that it is more pronounced for comparatively bad debtor countries. The coefficient on $BAD * FRAC * CAC$ is negative and highly significant. So apparently, creditors of stressed countries are well aware that not only the coordination embedded in their own bond contract, but also the composition of the rest of the debt matters in the event of a restructuring. This reflects the idea that holders of CAC bonds view other CAC bondholders as an asset as this will facilitate the likely restructuring, while they assess outstanding uncoordinated claims as a negative, since these bondholders might disrupt an efficient reorganization. Interestingly, the result for the aggregation effect is mostly driven by countries with a higher probability of a default event. In the set of *good* countries the effect is not significant.

Let us now analyze the coordination effect (propositions 2 and 5). We find that the coefficient of $BAD * FRAC$ is positive and highly significant, which contradicts proposition 5. However, as is argued in section 3.2, moral hazard can potentially overcome the positive coordination effect. So if moral hazard was much more a concern in the bad sample, this would indeed justify the results. Indeed we can imagine that the policy control over pledgeable output is of little interest for creditors in very good times. If the economy is doing great countries seldomly default, an economic rationale for which could be that the cost of default is very high in this times, e.g. proportional to output. In bad times however, dishonoring its external debt becomes a more lucrative option for debtors in financial straits. In so far we can judge the result as indicating, that creditors rightly view the merit of facilitated creditor coordination as a double-edged sword. For good creditors it is seen as an advantage to know that workouts will be orderly, while for bad creditors, the danger that the country misuses the clauses to walk away from its debt easily prevails. This is indeed the argument made by Eichengreen and Mody (2000) who find some evidence for this story in their analysis. The regression presented here supports this view.

Some empirical analysis of collective action clauses has asked whether coordination has been priced differently over time. For example it is argued that the governing law has been a disregarded dimension of sovereign bond contracts for a long time and only very recently has gained such prominence. To see whether the days of the pathbraking issue by Mexico in February 2003 and today's markets value coordination differently, I divide my sample in 2004-Q3. I define a dummy EARLY for quarters 2003-1 to 2004-3 and form interaction dummies similar to the exercise undertaken for countries with high spreads. The results are reported in the Appendix table 8. None of the coefficients is significant even at the 10% level. Consequently, I conclude that the pricing of coordination has not changed recently.

6.3 Existing Literature

As reviewed in the introduction most of the literature does not find evidence of any pricing peculiarities of bonds with collective action clauses. The findings just presented do not necessarily conflict with this idea. Generally speaking the analysis of Becker, Richards and Thaicharoen (2003) and Richards and Gugiatti (2003) of secondary market spreads are very similar to my regressions except for the inclusion of the variable FRAC as an explanatory variable. If FRAC and CAC∗FRAC are not included in the regression, the effect of CAC∗FRAC will partly show up in the coefficient for the collective action dummy. Intuitively, in my regression the variable CAC captures the effect the CAC inclusion has if the proportion of debt issued with CAC is 0% (which is only a logical exercise and of course not possible in the real world). In contrast, if CAC∗FRAC is dropped from the regression, CAC measures the effect of CAC inclusion for a country that has a proportion of debt issued with CAC that equals the sample average. So one would expect the effect of CAC inclusion to be downward biased if the composition of the debt is not used as an explanatory variable.

Eichengreen, Kletzer and Mody (2003) analyze primary market spreads and only include a variable similar to CAC∗FRAC in my investigation. They find insignificant results except for borrowers with very bad creditworthiness, who appear to pay higher spreads if their coordinated debt is higher. The pattern they observe (albeit not consistently significant) is similar to what I observe for the coordination effect, namely that a high share of coordination is viewed positively for good creditors and negatively for bad debtors. It may be that Eichengreen, Kletzer and Mody (2003) capture the effect more coordination has on all outstanding bonds as an effect it has on newly issued CAC bonds as they only include the interaction term CAC∗FRAC and not FRAC alone.

7 Conclusion

The paper presents a model to understand the pricing of bonds with and without clauses to facilitate creditor coordination in the secondary market for sovereign debt. The model has three main predictions: (1) Creditors should require a premium for holding bonds that include collective action clauses. (2) All creditors, irrespective of the contractual embodiment of the issue they hold, value improved coordination, as long as moral hazard is no overwhelming concern. (3) CAC bondholders value coordination more than other bondholders.

The paper also delivers an empirical test of the model. A dataset containing the sovereign bonds of 19 countries that matured after April 2003 is used for this matter. The data robustly supports the predictions of the model.

The main contribution of this findings is that the composition of the debt matters for the judgment of collective action clauses and that this judgment can not be undertaken by looking at each bond issue individually. This explains the difference between my findings and that of earlier studies on the pricing of collective action clauses. The degree of coordination, measured by the fraction of debt that contains provisions for facilitated creditor coordination, captures the expected disruption a restructuring would bring about. The effect of a variation in the degree of coordination can take either direction, depending on whether moral hazard is a concern. If moral hazard is negligible positive effects from increased efficiency in workouts prevail. Otherwise the creditors seek to implement a harsh punishment for any deviation of the contracted repayment. Then, collective action clauses are seen as a flaw.

Another important finding is that holders of coordinated bonds value the fact that they hold a CAC bond instead of an uncoordinated paper according to the fraction of coordinated bonds by the respective issuer outstanding. This suggests that CAC bondholders are aware that they, by allowing for facilitated restructuring, provide financial stability that is partly enjoyed by the holders of non-coordinated bonds as well. In a way, holders of uncoordinated bonds *free-ride* on the benignity of the CAC bond holders.

Appendix

explanatory variable	Coefficient	t-Statistic		Coefficient	t-Statistic	
Duration	0.2378	14.73	***	0.2799	14.76	***
$Duration^2$	-0.0125	-10.18	***	-0.0142	-9.96	***
CAC	0.1672	2.36	**	0.1524	1.79	*
FRAC	-2.0489	-6.52	***	-1.8057	-1.96	*
CAC∗FRAC	-0.9048	-4.69	***	-1.2312	-5.06	***
Euro	-0.0188	-0.53		-0.0249	-0.58	
Yen	0.3047	3.41	***	0.4087	4.08	***
£	-0.0670	-0.63		-0.0133	-0.11	
SF	0.7930	4.84	***	0.7927	4.69	***
2003 Q1	0.5310	7.44	***	0.2045	1.13	
2003 Q3	0.1471	2.49	**	0.0836	0.64	
2004 Q1	-0.0076	-0.14		-0.1166	-0.89	
2005 Q1	-0.0529	-1.01		-0.0852	-1.29	
2005 Q3	-0.4037	-7.39	***	-0.4303	-3.23	***
2006 Q1	-0.4921	-8.42	***	-0.5331	-3.87	***
GDP growth	n/a			-0.0264	-1.33	
Current Account	n/a			-0.0112	-0.35	
Inflation	n/a			0.0070	0.83	
log(GDP per capita)	n/a			1.5356	0.81	
Debt/GDP	n/a			7.7707	2.09	**
Debt/Exports	n/a			-0.1033	-0.76	
Debt/Reserves	n/a			0.1076	1.38	
Short-Term/Long-Term	n/a			0.3594	0.76	
log(Volume)	1.1380	7.17	***	1.1038	1.55	
on the run (6 month)	-0.1294	-2.22	**	-0.0924	-1.18	
Belize	9.2105	8.76	***	(dropped)		
Brazil	0.9108	17.40	***	0.8990	2.26	**
Chile	3.2366	6.02	***	2.6826	1.06	
Colombia	3.0337	9.69	***	3.1999	2.38	**
Costa Rica	4.5005	8.17	***	4.1489	1.69	*
Guatemala	5.8821	8.13	***	(dropped)		
Indonesia	6.3756	9.01	***	(dropped)		
Korea	3.4277	6.23	***	2.4316	0.79	
Panama	3.4141	8.14	***	(dropped)		
Peru	3.6165	8.73	***	4.3360	2.50	**
Philippines	3.2813	9.58	***	(dropped)		
Poland	1.2472	4.27	***	0.8397	0.63	
South Africa	2.5595	10.47	***	2.7685	2.75	***
Turkey	3.2638	6.75	***	3.1785	1.50	
Uruguay	5.9181	10.87	***	3.9689	1.53	
Venezuela	2.4921	14.36	***	(dropped)		
Constant	-8.6372	-4.80	***	-23.2185	-1.42	
Adjusted R^2	0.6452			0.6184		
Observations	1862			1282		

Table 6: Robustness Check: Regression 1

*,**,*** denote significance at the 10%,5% and 1% level respectively.

explanatory variable	Coefficient	t-Statistic		Coefficient	t-Statistic	
Duration	0.1727	12.59	***	0.2096	13.04	***
Duration^2	-0.0081	-7.78	***	-0.0099	-8.31	***
CAC	0.1534	2.00	**	0.0792	0.88	
FRAC	-1.1935	-4.19	***	-0.8711	-1.14	
CAC*FRAC	-0.0519	-0.24		-0.0188	-0.07	
BAD*CAC	-0.0372	-0.36		-0.0472	-0.38	
BAD*FRAC	0.7175	3.63	***	1.4212	4.49	***
BAD*CAC*FRAC	-1.3533	-4.30	***	-1.5770	-4.00	***
Euro	-0.0765	-2.56	**	-0.1082	-3.04	***
Yen	0.1032	1.38		0.1112	1.33	
£	-0.1954	-2.21	**	-0.1614	-1.68	*
SF	0.2598	1.88	*	0.2443	1.73	*
2003 Q1	0.5020	8.38	***	0.5475	3.66	***
2003 Q3	0.1470	2.98	***	0.3342	3.06	***
2004 Q1	0.0877	1.98	**	0.1639	1.50	
2005 Q1	0.0374	0.85		0.0265	0.49	
2005 Q3	-0.1401	-2.99	***	-0.3200	-2.90	***
2006 Q1	-0.1703	-3.33	***	-0.3857	-3.38	***
GDP growth	n/a			-0.0125	-0.76	
Current Account	n/a			-0.0149	-0.57	
Inflation	n/a			0.0020	0.29	
log(GDP per capita)	n/a			1.6242	1.02	
Debt/GDP	n/a			4.8607	1.58	
Debt/Exports	n/a			-0.2070	-1.85	*
Debt/Reserves	n/a			-0.1030	-1.58	
Short-Term/Long-Term	n/a			-0.5587	-1.43	
log(Volume)	0.4759	3.52	***	0.4109	0.70	
on the run (6 month)	-0.1379	-2.82	***	-0.0737	-1.13	
Belize	4.0199	4.47	***	(dropped)		
Brazil	0.3960	7.99	***	0.9916	3.00	***
Chile	1.1635	2.54	**	0.4000	0.19	
Colombia	1.2554	4.64	***	1.8920	1.70	*
Costa Rica	2.0397	4.34	***	1.6064	0.80	
Guatemala	2.6207	4.24	***	(dropped)		
Indonesia	2.9003	4.75	***	(dropped)		
Korea	1.1124	2.37	**	-0.3907	-0.15	
Panama	1.5705	4.39	***	(dropped)		
Peru	1.5165	4.26	***	2.4153	1.66	*
Philippines	1.2794	4.32	***	(dropped)		
Poland	0.1060	0.43		-0.4295	-0.39	
South Africa	1.3269	3.22	***	0.9729	0.56	
Turkey	1.1389	5.37	***	1.5412	1.83	*
Uruguay	2.8320	6.05	***	2.5621	1.20	
Venezuela	1.2320	8.07	***	(dropped)		
Constant (bad)	0.8816	16.72	***	0.8203	10.82	***
Constant	-1.1625	-0.76		-15.4053	-1.11	
Adjusted R^2	0.7539			0.7423		
Observations	1862			1282		

Table 7: Robustness Check: Regression 2

*,**,*** denote significance at the 10%,5% and 1% level respectively.

explanatory variable	Coefficient	t-Statistic	
Duration	0.2803	14.32	***
Duration^2	-0.0146	-9.91	***
CAC	0.2319	1.93	*
FRAC	-0.6865	-2.25	**
CAC∗FRAC	-1.2705	-3.93	***
EARLY∗CAC	-0.0307	-0.19	
EARLY∗FRAC	0.1797	0.46	
EARLY∗CAC∗FRAC	-0.2855	-0.50	
Euro	-0.0721	-1.63	
Yen	0.3285	3.17	***
£	-0.0887	-0.74	
SF	0.6595	3.77	***
2003 Q1	-0.0452	-0.35	
2003 Q3	-0.1720	-1.97	**
2004 Q1	-0.3684	-4.73	***
2005 Q1	-0.0845	-0.80	
2005 Q3	-0.2669	-2.40	**
2006 Q1	-0.3584	-3.06	***
GDP growth	-0.0150	-1.08	
Current Account	-0.0187	-1.49	
Inflation	0.0299	6.72	***
log(GDP per capita)	-0.8203	-4.87	***
Debt/GDP	1.6451	1.58	
Debt/Exports	0.1950	3.72	***
Debt/Reserves	0.1347	2.80	***
Short-Term/Long-Term	1.7221	3.96	***
log(Volume)	-0.0760	-3.29	***
on the run (6 month)	-0.0867	-1.04	
Constant	11.6697	7.43	***
Adjusted R^2	0.5880		
Observations	1282		

Table 8: Effects over time

*,** and *** denote significance at the 10%,5% and 1% level respectively.

References

Becker, T., A. Richards, and Y. Thaicharoen (2003): "Bond Restructuring and Moral Hazard: Are Collective Action Clauses Costly?," *Journal of International Economics*, 61(1), 127–161.

Chamberlin, M. M. (2002): "Sovereign Debt Contracts: What Do We Need to Change?," in *Can We Break the Crisis Cycle for Emerging Markets?*, London. Institute of International Finance Conference.

Corsetti, G., P. Pesenti, and N. Roubini (1999): "What caused the Asian currency and financial crisis?," *Japan and the World Economy*, 11(3), 305–373.

Dixon, L., and D. Wall (2000): "Collective Action Problems and Collective Action Clauses," *Financial Stability Review*, 8, 142–151.

Drage, J., and C. Hovaguimian (2004): "Collective Action Clauses: An Analysis of Provisions Included in Recent Sovereign Bond Issues," *Financial Stability Review*, 17.

Edwards, S. (1986): "The Pricing of Bonds and Bank Loans in International Markets : An Empirical Analysis of Developing Countries' Foreign Borrowing," *European Economic Review*, 30(3), 565–589.

Eichengreen, B., and A. Mody (2000): "Would Collective Action Clauses Raise Borrowing Costs?," Nber working papers, National Bureau of Economic Research, Inc.

——— (2001): "Would Collective Action Clauses Raise Borrowing Costs? An Update and Additional Results," International finance, EconWPA.

——— (2003): "Is Aggregation a Problem for Sovereign Debt Restructuring?," *American Economic Review*, 93(2), 80–84.

Eichengreen, B., and R. Portes (1996): *Crisis? What Crisis? Orderly Workouts for Sovereign Debtors.* CEPR, London.

Eichengreen, B. J., K. Kletzer, and A. Mody (2003): "Crisis Resolution: Next Steps," NBER working papers, National Bureau of Economic Research, Inc.

Elton, E. J., and T. C. Green (1998): "Tax and Liquidity Effects in Pricing Government Bonds," *Journal of Finance*, 53(5), 1533–1562.

Fernandez, R., and R. W. Rosenthal (1990): "Strategic Models of Sovereign-Debt Renegotiations," *Review of Economic Studies*, 57(3), 331–49.

FLEMING, M. J. (2003): "Measuring treasury market liquidity," *Economic Policy Review*, (Sep), 83–108.

IMF (2002): "Collective Action Clauses in Sovereign Bond Contracts - Encouraging Greater Use," Discussion paper, International Monetary Fund.

JANKOWITSCH, R., H. MSENBACHER, AND S. PICHLER (2006): "Measuring the liquidity impact on EMU government bond prices," *European Journal of Finance*, 12(2), 153–169.

KAMINSKY, G. L., AND C. M. REINHART (1999): "The Twin Crises: The Causes of Banking and Balance-of-Payments Problems," *American Economic Review*, 89(3), 473–500.

KRUEGER, A. O. (2002): "A New Approach to Sovereign Debt Restructuring," Discussion paper, International Monetary Fund.

OBSTFELD, M. (1986): "Rational and Self-fulfilling Balance-of-Payments Crises," *American Economic Review*, 76(1), 72–81.

——— (1996): "Models of currency crises with self-fulfilling features," *European Economic Review*, 40(3-5), 1037–1047.

RADELET, S., AND J. SACHS (1998): "The Onset of the East Asian Financial Crisis," *NBER Working Papers*, 6680.

RICHARDS, A., AND M. GUGIATTI (2003): "Do Collective Action Clauses Influence Bond Yields? New Evidence from Emerging Markets," *International Finance*, 6(3), 415–47.

SACHS, J. D. (1995): "Do we need an International Lender of Last Resort?," in *Frank Graham Lecture*. Princeton University.

TSATSARONIS, K. (1999): "The Effect of Collective Action Clauses on Sovereign Bond Spreads," *BIS Quarterly Review*, pp. 22–23.

WARGA, A. (1992): "Bond Returns, Liquidity, And Missing Data," *Journal of Financial and Quantitative Analysis*, 27(4).

WEINSCHELBAUM, F., AND J. WYNNE (2005): "Renegotiation, Collective Action Clauses and Sovereign Debt Markets," *Journal of International Economics*, 67(1), 47–72.

Dealing with country diversity: challenges for the IMF credit union model*

Gregor Irwin†, Adrian Penalver‡, Chris Salmon§ and Ashley Taylor¶

May 2008

Abstract

We develop a model in which countries can protect themselves against shocks by subscribing to a credit union that shares the key features of the International Monetary Fund, or by self-insuring through accumulating reserves. We assess the impact of the increasing heterogeneity of the Fund's membership on the political equilibrium Fund size and hence its effectiveness as a credit union. We find the Fund's existing lending framework is well suited to a world in which its members have homogeneous interests, but as the membership has become more heterogeneous the Fund is increasingly unlikely to provide financing on a sufficient scale to meet the demands of higher-risk members, leading them to rely more heavily on self-insurance. We conclude that the framework governing the Fund's lending operations may no longer be appropriate.

1 Introduction

The creation of the International Monetary Fund (IMF) in 1946 was a political solution to the economic challenge of ensuring international monetary co-operation. The IMF was placed at the apex of a monetary system based on fixed-but-adjustable exchange rate pegs, with responsibility for managing the system. Importantly, the IMF was provided with financial resources so that it

*The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. We are grateful to two anonymous referees for comments and suggestions and to participants at seminars at the Bank of England, the 2005 Annual Meeting of the Canadian Economic Association, the 2005 Annual Meeting of the Latin American and Caribbean Economic Association, the 2006 Congress of the European Economic Association, the 2006 Royal Economic Society Annual Conference 2006, the June 2007 FMG 'Cycles, Contagion and Crises Workshop' and Universitat Pompeu Fabra. This paper was finalised on 16 January 2008 and was published as Bank of England Working Paper No. 349, May 2008. JEL classification: F33, F34.

†Bank of England.

‡Bank of England.

§Bank of England.

¶Financial Markets Group, London School of Economics.

could ease the adjustment burden for countries experiencing temporary macroeconomic disequilibria. This lending role was necessary to ensure that countries had the appropriate incentives to eschew non co-operative behaviour and abide by the Fund's rulings.

Since 1946 the international monetary system, the Fund's oversight role in relation to it, and the composition of the Fund's membership have all changed markedly. The system of fixed-but-adjustable exchange rate pegs broke down in the 1970s. Since then floating exchange rate regimes have become much more widespread. The focus of the Fund has since shifted from managing the system to surveillance over the system, to ensure that policymakers take informed decisions, cognisant of the policy challenges faced by other countries and the economic linkages between countries. The membership of the Fund has more than quadrupled since its inception, expanding from a club of 44 industrialised countries in 1946, to become a near-universal institution with 185 members in 2007.

But during the same period the basic framework governing the Fund's lending operations has undergone much less change. It is still essentially a type of credit union into which countries pay a quota (or subscription) to become a member. Countries experiencing an adverse economic shock are entitled, under certain restrictions, to draw down their quota and temporarily borrow money from the Fund. The drawing (or access) right of each member is proportional to the size of its quota. Importantly, the overall size of the Fund, which determines how much crisis lending is available in aggregate, is voted on by the membership every five years.

This paper seeks to assess the efficiency of this credit union model, given the existing political decision-making process, and in the light of the changes in the Fund's membership that have occurred over its lifetime. We do this using a simple, yet novel theoretical model of the IMF as a credit union, in which the membership decides collectively by a vote on the size of the Fund and hence the amount of crisis lending it provides. This decision, in turn, impacts on individual country choices over the amount of self-insurance to hold in the form of reserves. The equilibrium Fund size and individual country reserve choices are analysed under three different characterisations of the decision-making processes – unconstrained majority voting, constrained majority voting, and qualified majority voting with an agenda setter. The welfare implications in each case are assessed and we consider how the existence of spillovers between countries affects the outcome.

In all cases the analysis suggests the current lending framework of the Fund may no longer be appropriate. It may well have been during the first two to three decades of the Fund's existence, when almost all countries were potential Fund debtors and had broadly homogenous interests, but the analysis suggests it is less well suited to the current situation in which members differ sharply in their economic characteristics and needs.[1] In particular, we find that with

[1] Between 1944 and 1977 industrialised and developing countries turned to the Fund for financial assistance. During that period five of the future G7 members borrowed from the

an increasingly heterogeneous membership, in terms of crisis probabilities, the decisions over the size of the Fund are likely to be driven by members with a relatively low crisis probability. Consequently, the Fund is increasingly unlikely to provide financing on a sufficient scale to meet the demands of higher-risk members, leading them to rely more heavily on self-insurance. The analysis suggests that increasing the size of the Fund may be Pareto improving, but only if the financial burden is distributed so that those who benefit most – that is, the countries which have the highest crisis probability – pay the most. This would constitute a significant change in the financing of the Fund's lending operations.

Our analysis is consistent with some of the more striking recent global economic developments, although clearly there are other potential explanations. First, it predicts that as the Fund's members become more diverse, then those countries most at risk of experiencing a payments imbalance will increasingly self-insure and hold more reserves. Second, it suggests that the development of intra-regional coinsurance mechanisms, such as the Chaing Mai initiative, are a natural artefact of the increasingly diverse interests of the Fund's membership, to the extent that the second-round spillover effects of crises are stronger within than across regions. Finally, the analysis is consistent with the trend increase in the average size of financial assistance granted by the Fund that has been observed over the past 30 years, and with the concerns about moral hazard that this trend has generated.

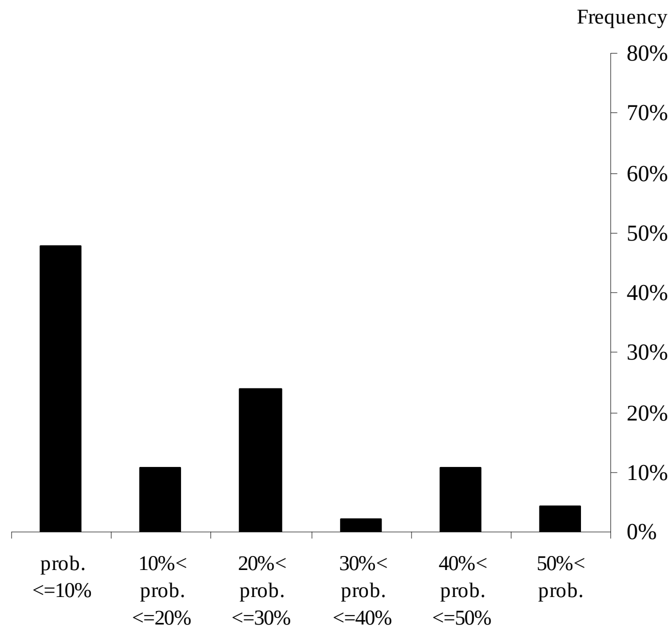
The increase in the diversity of the Fund membership is illustrated in Charts 1 and 2 which show how Fund lending activity has become increasingly concentrated on a (albeit still large) subset of the membership. This indicates that over the lifetime of the Fund the mean crisis probability of the membership has fallen, but the median crisis probability has fallen even faster. The distribution of crisis vulnerabilities has therefore become more skewed.

Chart 3 shows how the Fund has shrunk relative to GDP since its formation. In the first two decades since its foundation IMF members pledged quotas that amounted to around 1% of their GDP. In real terms the Fund has shrunk since then, so that by the end of 2005, a much larger Fund membership pledged quotas that amounted to around 0.7% of their GDP. Compared with world trade and capital flows the decline in quotas has been even starker, particularly during the past two decades. For example, Chart 3 illustrates that total IMF quotas have fallen from around 4%-5% of world merchandise trade (exports plus imports) in the period up to 1970 to 1.3% in 2005.

The increasing tendency towards self-insurance is quantified in Chart 4. As is well known the real value of reserve holdings has increased sharply over recent years and total reserves to IMF quota at end-2006 were three and a half times that at 1971, when the Bretton Woods system broke down. Most of that increase has taken place since the Asian crisis of 1997 and has been driven by countries in that region.

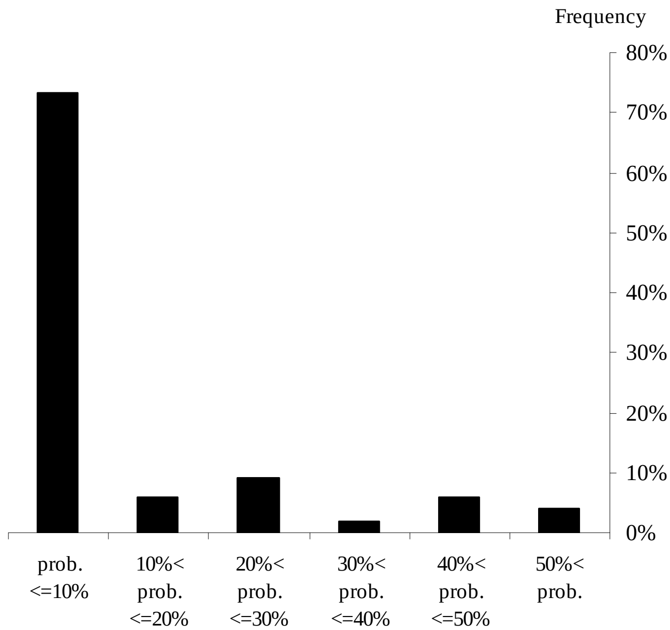
Fund, some repeatedly so. (The US and Germany are the exceptions.) But since 1977 the membership has become bifurcated.

Figure 1: Annual probabilities of borrowing from the Fund: 1951-65 (for members at end-1950)



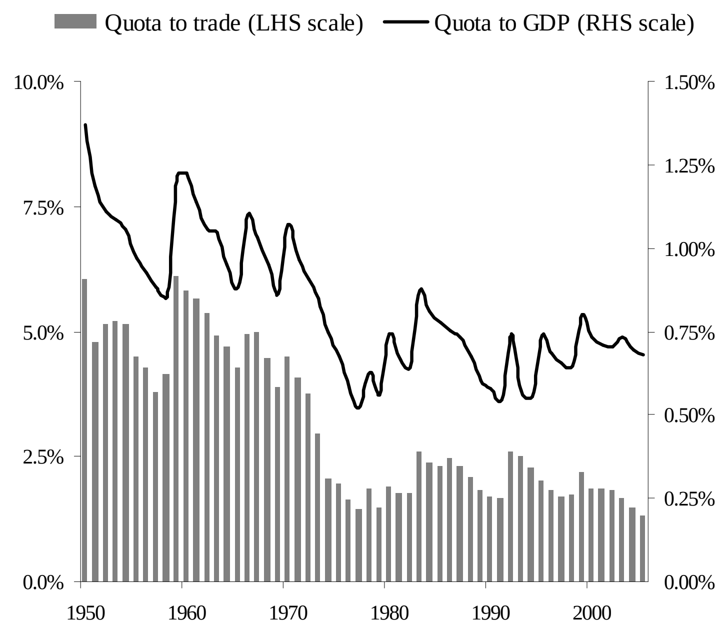
Sources: IMF *International Financial Statistics* (*IFS*) and authors' calculations.
Notes: Probability indicates the annual average probability of country making a purchase from the IMF General Resources Account, excluding reserve tranches, over the period. The sample of member countries is those for which *IFS* indicates non-zero quota at end-1950.

Figure 2: Annual probabilities of borrowing from the Fund: 1992-2006 (for members at end-1991)



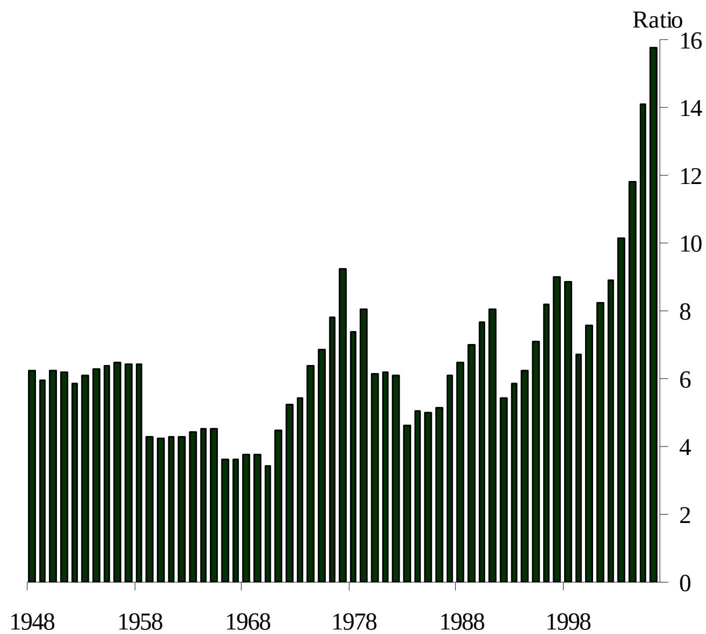
Sources: IMF *International Financial Statistics* (*IFS*) and authors' calculations.
Notes: Probability indicates the annual average probability of country making a purchase from the IMF General Resources Account, excluding reserve tranches, over the period. The sample of member countries is those for which *IFS* indicates non-zero quota at end-1950.

Figure 3: Average IMF quotas relative to member GDP and trade



Sources: IMF *International Financial Statistics* (*IFS*) and authors' calculations.
Notes: Each series is calculated as a weighted average (by denominator) for those members for whom both the quota and the denominator are available in each year. Trade is merchandise trade (sum of exports and imports).

Figure 4: Ratio of reserves to IMF quotas



Sources: IMF *International Financial Statistics* (*IFS*) and authors' calculations.
Note: The ratio is that of world reserves including gold to total IMF quotas.

Table 1: Borrowing from the Fund

	End-period quota share of borrowers:[1]	Average ratio of borrowing to quota[2]	
		all members:[3]	borrowers only:[4]
1948-1960	47%	1.7%	20.6%
1961-1970	44%	5.8%	29.7%
1971-1980	44%	6.9%	52.2%
1981-1990	36%	6.4%	57.2%
1991-2000	33%	6.4%	60.9%
2001-2006	17%	6.0%	172.9%

Sources: *IMF International Financial Statistics* (*IFS*) and authors' calculations.

(1) Borrowers defined as those making at least one purchase from the General Resources Account, excluding reserve tranches, during the period.

(2) Weighted by quota.

(3) Sum of purchases over sum of quotas across all country-year observations in each period.

(4) Sum of purchases over sum of quotas across all country-year observations in which purchases made during each period.

Finally, Table 1 shows how the average amount of borrowing from the Fund has changed over time. The weighted average ratio of borrowing to quota across the membership has remained stable over the past 40 years at around 6%. However over this period the fraction of the total IMF quota accounted for by those countries that borrow has shrunk from just under 50% to around 20%. Thus the average ratio of borrowing to quota for those countries that borrow has increased substantially. Changes to the distribution of crisis probabilities over this time, and in particular a decrease in the mean vulnerability, can help to explain this rise in 'exceptional access'.

The main message of this paper is that the framework governing the Fund's lending operations may no longer be appropriate. The existing credit union model was appropriate for a world in which the interests of the membership were homogenous. This may no longer be the case. The Fund comprises of increasingly heterogeneous countries. As a result, based on our model, the amount of crisis lending that is available from the Fund is likely to be suboptimally low, abstracting from concerns about moral hazard, increasing the reliance of members on economically inefficient self-insurance. An alternative approach may be needed: one which takes into account that creditor and debtor countries have different interests, but which also takes into account the moral hazard consequences of large-scale lending.

1.1 Modelling approach and related literature

The model employed to analyse a country's choice over Fund size is a simple one-period, partial equilibrium investment model. A country's demand for 'insurance' is motivated by the possibility that its final investment output may be reduced following a crisis with countries varying in their likelihood of suffering a crisis. As in reality, the insurance options available to each country include self-insurance via reserves and also subscription to an international credit union mechanism (the IMF) which entitles a country to access pooled resources in the event of a crisis.[2] Clearly in a world with a full set of contingent contracts there is no need for an external party such as the IMF to mitigate the costs of crises. However, despite the substantial development of international financial markets since the Fund's establishment, it remains the case that a range of market failures, such as inability to enforce sovereign debt claims across international borders, limit the ability of countries to insure against lower consumption states. Many developing countries continue to be excluded from world capital markets (either being quantity or price-rationed) particularly in crisis times when they most need the finance.

The analysis below draws on insights and approaches developed in a number of related literatures. For example, a clear analogy can be drawn with an individual's demand for private and public provision of health insurance. A country's crisis probability can be compared to the likelihood of an individual falling ill. Furthermore, an individual's choice of private insurance cover is often made conditional on the level of public insurance. So, in the model below, the political choice over the size of the credit union is taken before countries choose their level of self-insurance through reserve cover. In a similar manner Gouveia (1997) analyses the supplemental purchase of private health insurance above the level of public insurance and determines the political equilibrium level of provision of the latter by majority voting.

The political economy of risk-sharing across individuals in different countries via social insurance has been analysed in some detail in the context of fiscal federations. Motivated by developments within the European Union, recent papers, such as Persson and Tabellini (1996a, 1996b) and Alesina and Perotti (1998), examine the determination and characteristics of federal and state-level social insurance policies under various institutional arrangements. Many of the issues raised in such analyses, for example participation constraints on membership of the union, are of interest to our analysis. However, our focus is less on redistributive transfers and our level of analysis is the country rather than the region.

In this sense our approach is closely linked to recent political economy analysis of international organisations, in particular Alesina *et al* (2001, 2005). The focus of these papers is the provision of public goods with externalities by an international union. Clearly this differs somewhat from the credit union qualities of the IMF. Nevertheless the papers provide rich insights into the process of union formation, enlargement and decision-making (under both majority and qualified majority voting) which are in many ways applicable to the IMF.

While our modelling approach draws closely on the insights of the above literature, we believe it to be one of the first to model formally the political

[2] Cordella and Levy Yeyati (2006) provide a review of various other insurance-type mechanisms which, in theory, are potentially available to countries, for example capital controls, private insurance via contingent credit lines and regional swap arrangements.

economy of decision-making by the shareholders of the Fund and the trade-off faced by countries between self-insurance and IMF subscriptions.[3]

The structure of the rest of the paper is as follows. Section 2 sets out the basic one-period investment model which is the work-horse of our analysis. Section 3 derives the optimal self-insurance choices that countries would make in the absence of the Fund. Section 4 introduces the Fund into the model, and derives the size of the Fund preferred by individual members and the self-insurance choices that members make, contingent upon the Fund being a particular size. Section 5 considers the political equilibrium Fund size that arises under different assumptions, in particular regarding the voting process. Section 6 assesses the welfare implications of the political equilibrium. The final section draws together the conclusions.

2 Model set-up

The basic set-up is a one-period, partial equilibrium investment model in which returns are realised at the end of the period. Consumption then takes place. Demand for insurance is motivated by a country-specific potential for a crisis to hit immediately after the initial investment decision has been made. The key features are as follows.

2.1 Country characteristics

We consider N equally sized countries who only differ in their probability of a crisis. Country i has a crisis probability π^i drawn from a commonly known distribution with support $[\pi^1, \pi^N]$ with $0 < \pi^1 < \pi^2 < ... < \pi^N < 1$. Two key variables in our analysis are the median crisis probability, π^m, and the mean crisis probability, μ. The stylised facts demonstrate that, as the membership of the Fund has become more heterogeneous, π^m has fallen relative to μ. Below we argue that this has important implications.

Differing crisis probabilities are the minimum degree of heterogeneity required for our analysis. We are deliberately abstracting from many other issues such as economic size, economic structure or geopolitical significance to focus solely on the issue of relative risk. We also make a number of simplifying assumptions. First, the realisations of the idiosyncratic crisis risks are assumed to be independent. Second, we abstract from the question of moral hazard by assuming that the crisis probability is not affected by a country's policy effort or by its level of insurance. Third, unlike the analysis of, for example, Gouveia (1997) in relation to health insurance, we restrict the analysis to countries of the same income levels.

[3] For example, while Chami *et al* (2004) provide a model of IMF lending and postulate an objective function for the IMF they do not consider the optimal level of Fund subscriptions from a political perspective nor the interaction between Fund size and self-insurance.

2.2 Investment technology and insurance

Each country receives a unit endowment which it can invest in a project of type A, which yields an exogenous gross return of ρ_A if the country does not suffer a crisis. However, if a crisis occurs a proportion δ of this return is destroyed.[4] Countries can also place some of their endowment in technologies with insurance-like properties – self-insurance via reserves and/or access to payouts in the event of a crisis through subscription to the IMF. In this partial equilibrium model, and in common with related models of the international financial architecture, the returns on the various investment technologies are taken as exogenous and are assumed to be common across countries. Endogenising interest rates, for example through inclusion of a reserve asset in zero net supply, would not change the political analysis of Fund size choices in the sense that countries still take interest rates as given. The assumption of common returns across countries means that there is only one source of heterogeneity across countries, namely the probability of crisis, which facilitates the political economy analysis.

2.2.1 Reserves/self-insurance

If held through to the end of the period, reserves yield a certain return of ρ_R. However, in the event of a crisis, countries can switch their reserves into crisis investment projects. Assumption 1 below is sufficient to ensure that in a crisis all reserves are put into the new crisis investment project.[5]

Assumption 1: The gross return from the crisis investments, ρ_C, is greater than the simple reserve return, ρ_R.

One rationale for this assumption is that a crisis may lead to the destruction of the domestic capital stock (for example through the liquidation of capital stocks by foreign investors). This would increase the marginal product of capital, facilitating new investment opportunities with increased returns. An alternative interpretation is that reserves can yield a higher return than in the non-crisis state through mitigating the negative impact of a crisis on investment.

2.2.2 IMF crisis payouts

Membership of the IMF credit union has the advantage of allowing the country to access a greater potential pool of resources in the event of a crisis. Fund subscriptions (proportion g^i of the initial endowment) are repaid to members by the IMF with gross return ρ_F at the end of the period. For those countries hit by a crisis, a payout from the Fund of f^i is made immediately post-crisis which they must then pay back at the end of the period with gross interest rate ρ_F.[6] In reality the rate of remuneration paid by the Fund is lower than the rate

[4] As with project returns, we impose uniform crisis losses across countries. Allowing for δ^i varying by country would not add substantially to the analysis.

[5] In Section 3 the implications of Assumption 1 for relative consumption levels in crisis and non-crisis states are discussed.

[6] We thus assume that states of nature are verifiable in determining the payout and that there are no contract enforcement problems in ensuring repayment. Furthermore, we assume

at which IMF lending is repaid by crisis economies, with the wedge between the two helping to finance Fund expenses.[7] Adding an exogenous wedge between these two rates within the model would provide little additional insight. In terms of the relation between ρ_F and ρ_R, in the model they are both exogenous and unrelated although in reality they are linked in the sense that the IMF interest rates are related to those on widely held reserve assets. The IMF's basic rate of remuneration and rate of charge are based on the Special Drawing Rate (SDR) interest rate. This interest rate is a weighted average of short-term money market interest rates, namely the Eurepo rate and UK, Japanese and US short-term government bills. These assets, or similar longer-term securities, can be held as reserve assets providing a linkage between ρ_F and ρ_R.

Assumption 2 below is sufficient to ensure that in a crisis all payouts received from the Fund are put into the new crisis investment project.

Assumption 2: The gross return from the crisis investments, ρ_C, is greater than the return paid on Fund subscriptions, ρ_F.

The Fund payout is pinned down by the chosen subscription levels through the Fund's budget constraint. To keep the analysis tractable, as in the health insurance analysis of Gouveia (1997), we employ the expected budget constraint rather than employing the budget constraints for all realisations of nature. If the *ex-post* budget constraint is employed then an additional stage of voting on *ex-post* subscription increases would be required. Furthermore, in the long term, which is the focus of the static model, the Fund's expected budget constraint is likely to hold (ie the expected total crisis payouts equal the size of the Fund).[8] With country-specific subscription levels and payouts, the expected budget constraint is $\sum_{i\in\mathcal{H}} \pi^i f^i = \sum_{i\in\mathcal{H}} g^i$ where $\mathcal{H}$ is the set of H countries that are members of the Fund. With common subscription levels and payouts this simplifies to $\mu f = g$, where μ is the mean crisis probability of those countries within the Fund. Thus, for a given subscription level, the higher is μ, the lower the crisis payouts are. While this does not fit with the Fund's formal access limits relative to quota, it does seem broadly consistent with trends of actual disbursements over time. Since the 1970s the proportion of members (by quota share) accessing Fund resources has fallen, as illustrated in Charts 1 and 2, which can be thought of as representing a fall in the mean crisis probability. At the same time those borrowing have accessed increasing funds relative to quota, as illustrated in Table A.

---

that the payout is automatic. In reality, access to IMF resources above a member's reserve tranche has to be approved post-crisis. However this would require a second round of voting within the model. In addition, in practice, excluding exceptional access cases, access to resources beyond the reserve tranche for crisis economies can usually be assumed to be forthcoming, albeit with conditionalities.

[7] For example, for the week 7-13 May 2007, the adjusted rate of remuneration was 4.06% while the adjusted rate of charge was 5.5% (see www.imf.org/external/np/tre/sdr/burden/2007/050707.htm). The former is the interest rate on repayment to members on their remunerated reserve tranche while the latter is the interest rate on a member's outstanding credit to the Fund.

[8] For sufficiently large numbers of countries this assumption can also be justified through appeal to the law of large numbers.

2.3 Preferences

Countries maximise their expected utility from consumption at the end of the period:

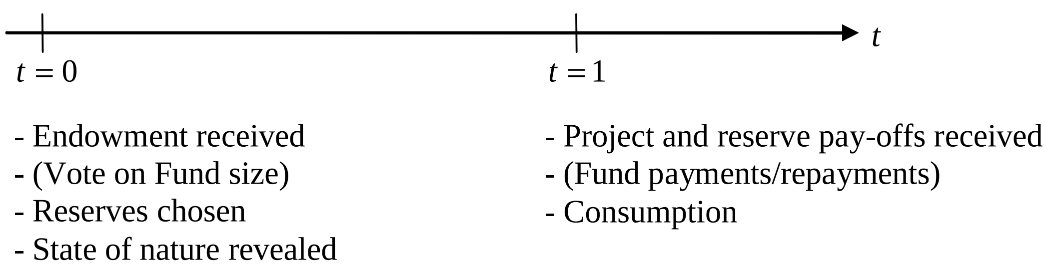
$$W^i = \pi^i u(c_c) + (1 - \pi^i) u(c_n) \tag{1}$$

where c_c is consumption in the crisis state and c_n is consumption in the non-crisis state. The state-specific utility functions are event independent with $u'(c) > 0$, $u''(c) < 0$ and standard Inada conditions $\lim_{c \to \infty} u'(c) = 0$ and $\lim_{c \to 0} u'(c) = \infty$.

It could be that countries also care about outcomes in other countries for a variety of reasons. For example, crises overseas may spill over to home consumption via trade and financial flows, or there may be concerns for others' consumption for political or altruistic reasons. For simplicity, we ignore possible spillover effects in the main body of this paper. However, in Appendix A we extend the analysis to incorporate spillover effects and assess their likely impact. In particular, we demonstrate that if a high weight is placed on spillovers from countries with high crisis probabilities, then this is likely to result in a larger Fund size, other things being equal, as lower crisis probability countries perceive that IMF crisis payouts offer greater benefits.

To recap, Chart 5 provides a summary timeline of the model. As an initial reference point, in the following section we derive a country's optimal reserve choice in a world with no Fund.

Figure 5: Timeline (features present with Fund indicated in parentheses)



3 World with no Fund

With no Fund (denoted ^{nf}), country i chooses reserves, $b^{i,nf}$, and the residual level of investment in project A, equal to $1 - b^{i,nf}$, to maximise expected utility subject to the constraint that reserves are between zero and one. The consumption level in the crisis state is:

$$c_c^{i,nf} = \rho_A (1 - \delta)(1 - b^{i,nf}) + \rho_C b^{i,nf}$$

In the crisis state, loss-adjusted returns of $\rho_A(1-\delta)$ are earned on the investment in project A, with returns of ρ_C earned on the funds placed in reserves which are then transferred to the crisis investment project. In the non-crisis state, consumption is

$$c_n^{i,nf} = \rho_A(1 - b^{i,nf}) + \rho_R b^{i,nf}$$

In this state the full return of ρ_A is earned on the investment in project A and ρ_R is earned on reserve holdings.[9] The constrained optimisation problem for country i is the following Lagrangean:

$$\underset{b^{i,nf}}{max}\mathcal{L} = \pi^i u(c_c^{i,nf}) + (1 - \pi^i)u(c_n^{i,nf}) + \theta_1 b^{i,nf} - \theta_2(b^{i,nf} - 1) \qquad (2)$$

Assumption 3 below is required to ensure that the crisis loss incurred on the return on the initial investment project, δ, is high enough so that some countries choose to invest in reserves.[10]

Assumption 3: $\delta > 1 - \rho_C/\rho_A$

Proposition 1 *In a world with no Fund, optimal reserve holdings are increasing in a country's crisis probability. Countries with crisis probability below* $\underline{\pi}^{b,nf}$ *hold zero reserves while countries with crisis probability above* $\bar{\pi}^{b,nf}$ *put all their endowment in reserves.*

Proof. See Appendix B.1. ■

As expected, under Proposition 1, the higher a country's crisis probability, the higher the insurance via reserves. This is illustrated in Chart 6. The level of the reserves is also non-decreasing in the severity of the crisis, captured by δ, and depends on the returns on both the investment projects and reserves.[11]

4 Country choice of optimal Fund size and reserves

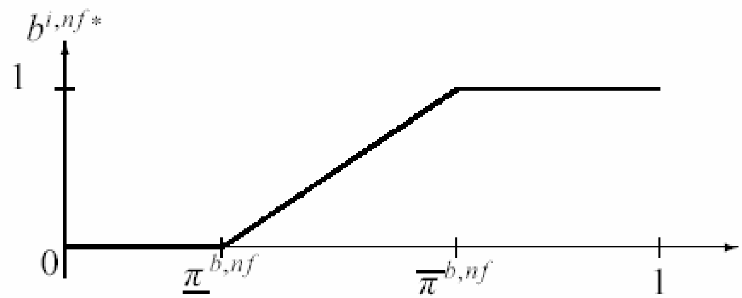
We now analyse how countries choose their overall level of insurance when both reserves and Fund membership are available (denoted ^{f}). As countries are

[9] The assumption that $\rho_C > \rho_R$ could potentially give rise to a country's consumption crisis level being higher than in the non-crisis state. This is, however, only the case if the crisis probability is high enough (for log utility we require the condition $\pi^i > \frac{(\rho_A - \rho_R)}{\delta\rho_A + (\rho_C - \rho_R)}$).

[10] This assumption is required to obtain positive interior values for reserve holdings (see Appendix B.1).

[11] The sensitivity of the optimal reserve choice to these returns depends on the concavity of the utility function. With log utility the level of reserves increases with ρ_R and ρ_C and falls with ρ_A. As would be expected, with a constant relative risk aversion (CRRA) utility function, the level of reserves also varies with the coefficient of relative risk aversion. Interestingly, the sensitivity of reserves to the degree of relative risk aversion depends on the level of the crisis probability. For small (large) crisis probabilities the level of reserves rises (falls) with the level of risk aversion. The intuition is that if the crisis probability is sufficiently high then, for given reserves, the higher risk aversion reduces overall marginal expected utility with respect to reserves, as it has a greater negative effect on expected marginal utility in the crisis state than the positive effect on marginal utility in the non-crisis state. Thus, for such crisis probabilities, higher risk aversion leads to falling optimal reserves.

Figure 6: Optimal choice of reserves with no Fund



assumed to be identical in size, and given the present institutional arrangement at the Fund which links quotas to measures of economic size, we consider the case in which there are common subscription levels and common payouts in the event of a crisis for all Fund members. Reserves supplement the crisis payouts available from membership of the Fund.[12] In this section we derive two key elements of the solution to our problem, before determining the political equilibrium provision of Fund coinsurance in the next section. These elements are (1) the amount of self-insurance that each country will choose, contingent upon a given Fund size, and (2) the Fund size that each prefers, given that quotas are allocated uniformly to all members.

4.1 Country choice of reserves for given Fund size

For a given Fund size, countries must decide whether to supplement their crisis consumption insurance from IMF membership with additional reserves. Consumption in the non-crisis state differs from the no-Fund world through the effect of the initial Fund subscription g which receives a return ρ_F. With common Fund subscription rates, non-crisis consumption is given by

$$\begin{aligned} c_n^{i,f} &= \rho_A(1 - b^{i,f*}(g) - g) + \rho_R b^{i,f*}(g) + \rho_F g \\ &= \rho_A - (\rho_A - \rho_R) b^{i,f*}(g) - (\rho_A - \rho_F) g \end{aligned}$$

In the crisis state, the Fund payout can be invested in the crisis investment

[12] The situation is similar to the problem analysed by Crémer and Palfrey (2000) and Alesina *et al* (2005) in the context of federal public goods provision, in which decisions over country policy are made following a decision on provision at the federal level.

technology, but must be paid back at return ρ_F, so crisis consumption is

$$\begin{aligned} c_c^{i,f} &= (1-\delta)\rho_A(1-b^{i,f*}(g)-g)+\rho_C b^{i,f*}(g)+((\rho_C-\rho_F)/\mu+\rho_F)g \\ &= (1-\delta)\rho_A+\Gamma b^{i,f*}(g)+\Omega g \end{aligned}$$

where $\Gamma \equiv \rho_C - (1-\delta)\,\rho_A$ and $\Omega \equiv (\rho_C-\rho_F)/\mu+\rho_F-(1-\delta)\,\rho_A$.

Taking the Fund subscription level, g, as given, the constrained optimisation problem for country i is:

$$\max_{b^{i,f}} \mathcal{L} = \pi^i u(c_c^{i,f}) + (1-\pi^i)u(c_n^{i,f}) + \lambda_1 b^{i,f} - \lambda_3(b^{i,f}+g-1) \quad (3)$$

Focusing on the non-trivial case in which $0 < g < 1$, countries choose to supplement their Fund insurance with reserves if their crisis probability is high enough.

Proposition 2 *For $0 < g < 1$, countries will have positive additional reserve holdings if their crisis probability lies in the range $[\underline{\pi}^{b,f}, \bar{\pi}^{b,f}]$. The preferred level of reserves is increasing in the crisis probability and decreasing in the size of the Fund, g. For $\pi^i < \underline{\pi}^{b,f}$ no reserves are held, while for $\pi^i > \underline{\pi}^{b,f}$ all the endowment is put into the Fund and reserves.*

Proof. See Appendix B.2. ■

The cut-offs $\underline{\pi}^{b,f}$ and $\bar{\pi}^{b,f}$ depend on the returns on the investment projects and reserves and on the value of g. This proposition illustrates the substitution in the form of insurance which takes place as the Fund size increases. As g rises the level of supplemental reserves falls and the threshold for the crisis probability at which countries add further reserves increases.[13] This result is similar to that of Gouveia (1997) in which higher public provision of health insurance may reach 'choking levels', crowding out private provision.

4.2 Country preferences over Fund size

We now consider individual country preferences over the size of the Fund and hence the uniform subscription level for all members. Country i derives its policy preference over g^i knowing how it would augment reserves in the second stage. Thus it chooses g^i by solving the following constrained optimisation problem:

$$\begin{aligned} \max_{g^i} \mathcal{L} = \pi^i u(c_c^{i,f}(g^i, b^{i,f*}(g^i))) + (1-\pi^i)u(c_n^{i,f}(g^i, b^{i,f*}(g^i))) \\ + \lambda_1 b^{i,f*}(g^i) + \lambda_2 g^i - \lambda_3(b^{i,f*}(g^i)+g^i-1) \quad (4) \end{aligned}$$

Define $\tilde{\mu} \equiv \{1 + \frac{(\rho_C-(1-\delta)\rho_A)(\rho_R-\rho_F)}{(\rho_C-\rho_F)(\rho_A-\rho_R)}\}^{-1}$.

[13] With CRRA preferences this crowding out is very apparent – the optimal reserve choice with the Fund is equal to the reserve choice in a no-Fund world minus an adjustment in proportion to the size of the Fund.

Proposition 3 *Individual country preference over Fund size:*

(a) For $\mu < \tilde{\mu}$, all countries prefer a Fund size that is non-decreasing in their crisis probability, with interior solutions in the range $[\underline{\pi}^{g,f}, \bar{\pi}^{g,f}]$. The optimal supplemental reserve holding at the preferred Fund size is zero.

(b) For $\mu > \tilde{\mu}$, all countries prefer no Fund. The optimal reserve holdings are determined by Proposition 1.

(c) If $\mu = \tilde{\mu}$, all countries prefer either no Fund and zero reserves or a positive Fund size and positive reserves, with the choice increasing in the crisis probability.

Proof. See Appendix B.3. ■

Recall from the Fund's budget constraint that $\mu = g/f$, so $\tilde{\mu}$ can be viewed as a threshold value of the ratio of the initial subscription to the crisis payout. Under Proposition 3, provided the mean crisis probability is sufficiently low ($\mu < \tilde{\mu}$), each country would prefer to hold zero reserves and use the IMF to provide additional funds in the event of a crisis. This is because f is sufficiently greater than g so that the member gets high leverage out of the initial subscription. If the mean crisis probability is too high ($\mu > \tilde{\mu}$), each country would prefer a zero Fund size and the only insurance would be self-insurance. The intuition is that a high mean crisis probability reduces the gross expected utility gain from a given Fund subscription, since more countries are likely to share the fixed total pot for payouts. Note that $\tilde{\mu} \geq 1$ if $\rho_F \geq \rho_R$.

We focus on part (a) of Proposition 3 since our primary interest is in a world with a positive Fund size. With an interior solution under part (a), the level of Fund subscription $g^{i,f*}$ preferred by country i is increasing in that country's crisis probability and in the severity of the crisis (ie the size of δ).[14]

5 Political equilibrium

We have now determined that Fund members will have different preferences over g and in particular those with a higher vulnerability will tend to prefer a larger Fund size. How, then, is the actual Fund size determined? In this section we focus on the case where $\mu < \tilde{\mu}$ and consider whether there is a political equilibrium outcome when the size of the Fund is determined by a vote of the Fund's membership.

Under the Fund's Articles and Agreements (Article XII, Section XII), the vote allocation of each member for decisions of the Fund's Board of Governors or

[14] In general, the sensitivity of the Fund size to the other parameters is dependent upon the concavity of the utility function. With CRRA preferences again the choice of Fund size varies with the degree of risk aversion. As mentioned in relation to the reserve choice in the no-Fund world, the sign of this relationship depends on the crisis probability. For π^i low (high) enough the optimal Fund size increases (decreases) with the degree of risk aversion. For log utility, the optimal choice of Fund is increasing in ρ_C and decreasing in ρ_A. The optimal choice of Fund size falls with μ (for the reasons discussed above). The sensitivity of the preferred Fund size to ρ_F is ambiguous, depending on π^i. On the one hand a higher ρ_F increases consumption in the non-crisis state but, on the other hand, it increases repayments in the crisis state.

Executive Board is equal to 250 'basic votes' plus an extra vote for each 100,000 special drawing rights of its quota.[15] Thus, for larger countries the voting share is slightly less than the quota share while for smaller countries the voting share is above the quota share. Nevertheless, it is a reasonable approximation to equate quotas with voting shares.

As noted in the introduction, quotas – and hence, votes – reflect both economic size and openness, with the US holding the largest country quota. Allowing for bloc voting by larger members complicate the analysis of the political equilibrium and generally does not provide significant additional insights.[16] Consequently, unless otherwise stated, we make the simplifying assumptions of equal country size and hence equal voting shares.

In the following subsection we consider the outcome when there is unconstrained majority voting, which allows us to employ the median-voter theorem to solve for a political equilibrium Fund size. In reality there may be a binding constraint on the decisions taken by a median voter. For example, it is possible that low crisis probability members may prefer to leave a Fund which they regard as being too large. Consequently, in Subsection 5.2 we consider when this participation constraint is likely to bind and the possible implications of this for the majority-voting equilibrium. Finally, although general decisions of the Fund are based on majority voting, the more important decisions actually require a qualified majority. For example, a revision to quotas (which is the policy choice variable in our model) requires an 85% majority. In the final subsection we consider how decisions might be taken under this scenario, assuming there is an agenda setter who determines the choices that are put to the vote by the membership of the Fund.

5.1 Unconstrained majority voting

Following the political economics literature, in order to generate equilibrium policies through pure majority rule, restrictions must be made on either the form of preferences over policy or the institutional framework.[17] In this subsection we take the former approach and check that the required conditions on preferences are satisfied. In doing so we follow a similar approach to the aforementioned literature on international political unions (see, for example, Alesina *et al* (2001, 2005)) and risk-sharing in federations (see, for example, Persson and Tabellini (1996a, 1996b)).

First we consider the case where all countries must be in the Fund and where they face a uniform subscription level and payout in the event of a crisis. The latter assumption reflects the current reality and adds tractability. Denote by π^m the crisis probability of the median country in the entire set of N countries.

With our single policy variable of the Fund subscription level, preferences

[15] The number of basic votes is currently under review.

[16] At the margin bloc voting is likely to give greater influence to the country whose bloc of votes straddles the critical threshold for either a majortity vote or a qualified-majority vote.

[17] Pure majority rule is characterised by direct democracy, sincere voting and an open agenda.

exhibiting single-peakedness or the single-crossing property can generate a political equilibrium under pure majority rule (ie are sufficient for the median voter theorem to hold).[18] As in Persson and Tabellini (1996b), the median voter theorem applies since the only source of heterogeneity among voters is the probability of a bad outcome (in our case a crisis, in their paper unemployment) which enters into preferences in a linear manner.

Proposition 4 *Given $\mu < \tilde{\mu}$ and with countries voting over the size of a Fund to be applied uniformly to the entire set of N countries, the political equilibrium Fund size is determined by the median voter theorem and is the optimal choice of the country with the median crisis probability. This optimal choice, $g^*(\pi^m)$, is as defined in Proposition 3(a).*

Proof. See Appendix B.4. ■

Thus, if $\pi^m \in [\underline{\pi}^{g,f}, \bar{\pi}^{g,f}]$ we have a positive Fund size which is increasing in the median country's crisis probability.[19]

Putting together Propositions 2 and 3 enables comparison of the levels of investment with and without the Fund. For $\mu < \tilde{\mu}$ and with a positive Fund size countries will hold supplemental reserve holdings if they have sufficiently high crisis probability $\pi^i > \underline{\pi}^{b,f}$. Adding Proposition 1 enables us to compare the crisis probability at which a country will begin to self-insure via reserves in the world with and without the Fund. Chart 7 illustrates the possible relative cut-off points for reserve holdings between the no-Fund and Fund worlds.[20]

Chart 8 provides a graphical example of the level of residual investment.[21] While investment is higher for lower crisis probability countries in the no-Fund world (as they are not forced to insure through a Fund subscription), countries begin to self-insure at a lower crisis probability in the no-Fund world than in the Fund world (ie $\underline{\pi}^{b,f} > \underline{\pi}^{b,nf}$). Similarly they reach the respective corner solutions for reserve holdings at a lower crisis probability in the no-Fund world ($\bar{\pi}^{b,f} > \bar{\pi}^{b,nf}$).

The key implication of this analysis is that it is π^m and not μ which drives the equilibrium Fund size. Therefore, if the distribution of crisis probabilities becomes more skewed, so that π^m falls relative to μ, as is consistent with the stylised facts, then we would expect that the Fund will decrease in size and so provide less coinsurance.

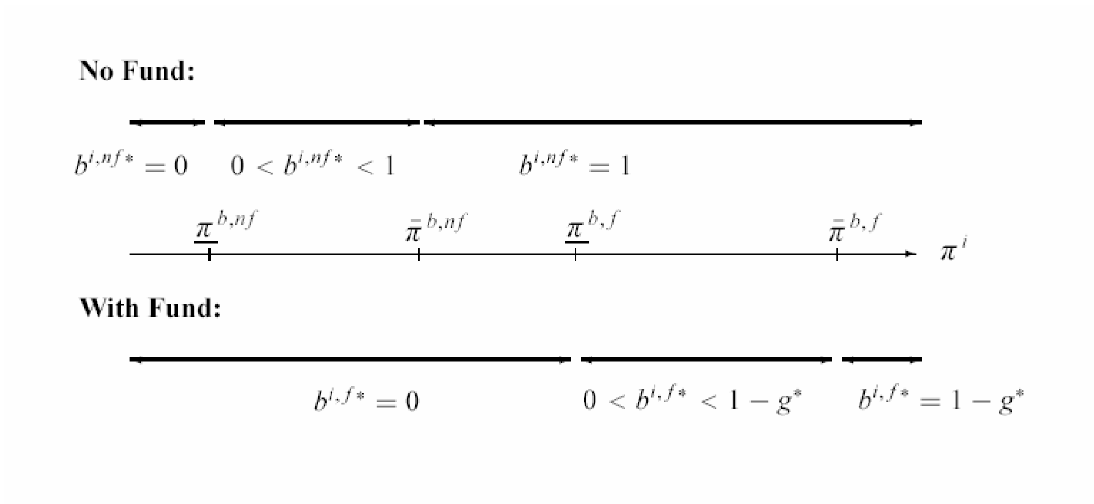
[18] See Gans and Smart (1996) and Persson and Tabellini (2000). As detailed in Persson and Tabellini (2000, pages 22-23), with single-peaked preferences over different policy options, a Condorcet winner always exists (ie a policy which beats any other feasible policy in a pairwise vote) and is the median-ranked preferred policy. If all preferences exhibit the single-crossing condition then the policy preferences of voters can be ordered by their types. In this case again a Condorcet winner always exists, but is the preferred policy of the median-ranked individual by type.

[19] For $\pi^m < \underline{\pi}^{g,f}$, the political equilibrium Fund size is zero, while for $\pi^m > \bar{\pi}^{g,f}$ the political equilibrium Fund size is one.

[20] This is one of two possible rankings of the cut-off points. It is ambiguous whether $\bar{\pi}^{b,nf} \gtrless \underline{\pi}^{b,f}$.

[21] We assume a constant relative risk aversion utility function with coefficient of relative risk aversion equal to 3, $\rho_C = 1.075$, $\rho_A = 1.05$, $\rho_R = \rho_F = 1.025$, $\mu = 0.025$, $\delta = 0.1$.

Figure 7: Optimal reserve holdings by crisis probability



5.2 Constrained majority voting

So far we have assumed that all countries are members of a Fund whose subscription level is determined by the country with the overall median crisis probability. For simplicity we have ignored the potential for participation constraints to bind on Fund membership.[22] However, assuming that redistributive transfers between countries are not feasible, it may well be the case that for sufficiently high subscription levels some countries would be better off leaving the Fund.[23] Clearly this would affect the composition of the Fund's membership and hence the equilibrium size of the Fund.

Denote the expected utility difference between being in a Fund of positive size and outside the Fund as $D^i(g, \pi^i) \equiv W^i(0 < g < 1, \pi^i) - W^i(g = 0, \pi^i)$.

Proposition 5 *The expected utility difference between being in a Fund of given positive subscription level g and outside the Fund is increasing in a country's crisis probability $\partial D^i / \partial \pi^i > 0$. Any country with crisis probability π such that $0 < \pi < \hat{\pi}(g)$ strictly prefers not to be in the Fund.*

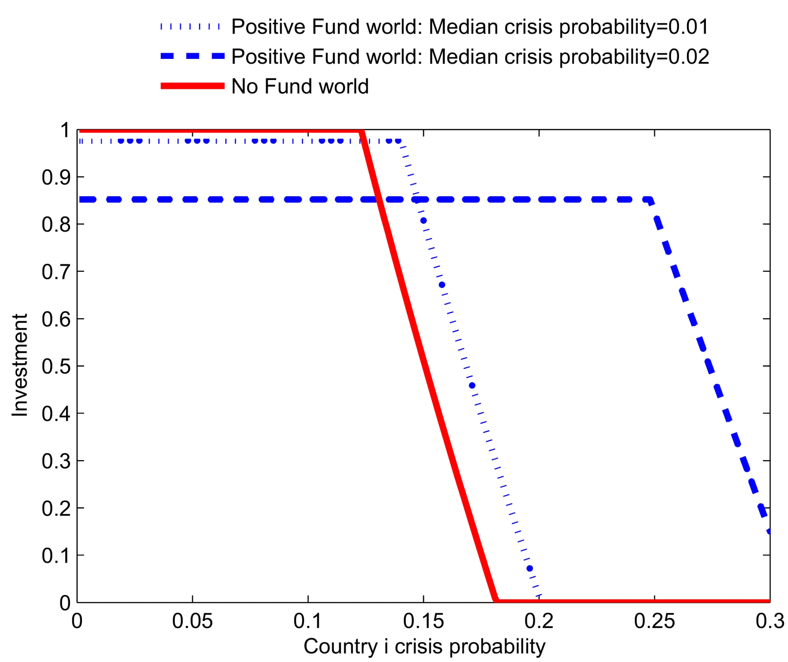
Proof. See Appendix B.5. ■

Proposition 5 implies that, for a given Fund subscription level of g, if the lowest crisis probability in a Fund of all N countries, π^1, is low enough ($\pi^1 <$

[22] Such constraints are discussed in detail for international unions in Alesina *et al* (2001, 2005).

[23] Our game-theoretic focus is on the possibility of withdrawal from the Fund, rather than on entry to the Fund in the first place. This is because Fund membership is open to any country that satisfies the terms prescribed by the Board of Governors (which should be consistent with principles applied to existing members). Thus there is no formal vote on new membership.

Figure 8: Optimal investment with and without a Fund



Notes: Model with constant relative risk aversion utility. Parameter values: ρ_A = 1.05, ρ_C = 1.075, ρ_R =ρ_F = 1.025, μ = 0.025 and coefficient of relative risk aversion = 3.

$\hat{\pi}(g))$ then there is at least one country who prefers not to be in such a Fund.[24] This is more likely to occur the greater is the difference between π^1 and π^m.

The key question is how does the median voter react to this participation constraint? The median voter has two options: either reduce g so as to keep the lower-bound member in the Fund, or allow the lower-bound member to drop out. In the latter case the putative median voter may lose its privileged position and be replaced by a new median voter as the size of the membership decreases. Which option is preferred is likely to depend on the shape of the distribution function for crisis probabilities in general, and in particular the crisis probability of the putative median voter relative to that of both the lower-bound member and the new median voter if countries drop out of the Fund. Moreover, depending on the distribution of country crisis probabilities, the dropping out process could continue until the Fund unravels and the Fund ceases to exist.[25] It is not possible to pin down whether this will be the case or, if not, what the equilibrium number of countries in the Fund will be, without specifying the exact distribution function for crisis probabilities and model parameters. However, we can characterise the nature of subscription levels for an equilibrium Fund with a stable number of members.

Suppose that the Fund initially has $\mathcal{M}$ members whose crisis probabilities, under Proposition 5, are ranked from π^{N-M+1} to π^N. The median country has crisis probability $\pi^m(\mathcal{M})$ and if unconstrained would set the Fund subscription level at $g^u(\pi^m(\mathcal{M}))$. If the participation constraint binds, ie $\pi^{N-M+1} < \hat{\pi}(g^u(\pi^m(\mathcal{M})))$, then the median voter could reduce the Fund subscription level to its constrained value of $g^c(\pi_{N-M})$ such that lower-bound member is indifferent between remaining in and leaving the Fund. Alternatively the putative median voter could allow those lower-bound members who would prefer not to be in a Fund of subscription level $g^u(\pi^m(\mathcal{M}))$ to drop out. This would result in a new Fund with median crisis probability country $\pi^m(\mathcal{R})$, where $\mathcal{R} < \mathcal{M}$ and $\pi^m(\mathcal{R}) > \pi^m(N)$, who would then go through the same process as above.

Given that, as a matter of fact, no low-crisis probability country has ever withdrawn from the Fund, it is interesting to consider under what conditions a stable Fund of $\mathcal{M}$ members is likely to exist. For $\frac{\partial\hat{\pi}(g^u(\pi^i))}{\partial\pi^i} > 0$ it can be shown that if the median country $\pi^m(\mathcal{M})$ of a Fund of $\mathcal{M}$ members has a crisis probability above a cut-off level (denoted by $\breve{\pi}(\mathcal{M})$) it will face a constrained choice of Fund subscription. With a standard log utility function it can be shown that $\frac{\partial\hat{\pi}(g^u(\pi^i))}{\partial\pi^i} > 0$.[26] Under this condition, an equilibrium Fund of $\mathcal{M}$

[24] We are implicitly assuming that there can be at most one Fund (ie we discount for now the possibility that countries at higher crisis probabilities may wish to leave the existing Fund to set up their own Fund with higher subscription levels).

[25] Concerns over participation constraints are also raised in Persson and Tabellini (1996b). They consider a second stage ratification vote which is imposed after the choice of federal policy which leads to repeal of the federal arrangement if either country rejects. In this case if there are large divergences in regional risk across the federation then the threat of secession leads to an upper bound on the level of federal insurance.

[26] Since $\frac{\partial g^u(\pi^i)}{\partial\pi^i} > 0$ the condition $\frac{\partial\hat{\pi}(g^u(\pi^i))}{\partial\pi^i} > 0$ is equivalent to $\frac{\partial\hat{\pi}(g^u(\pi^i))}{\partial g^u} > 0$. With a log utility function this can be shown to hold for all $\hat{\pi}(g(\pi^i)) < \pi^i$. Note that if $\hat{\pi}(g(\pi^i))$ is

members can be characterised as follows:

- If $\pi^m(\mathcal{M}) < \breve{\pi}(\mathcal{M})$ then the median country's unconstrained choice of Fund subscription is the majority voting equilibrium. All $\mathcal{M}$ countries prefer to be in a Fund with subscription level $g^u(\pi^m(\mathcal{M}))$ than to be outside the Fund.

- If $\pi^m(\mathcal{M}) > \breve{\pi}(\mathcal{M})$ then, if $\mathcal{M}$ members are in the Fund in equilibrium, the only possible Fund subscription level is the constrained choice which satisfies $D^{N-M+1}(g^c(\pi^m(\mathcal{M})), \pi^{N-M+1}) = 0$. All $\mathcal{M}$ countries weakly prefer to be in the Fund.

Thus the observation that no low crisis probability countries have dropped out of the Fund is consistent with the model if there is either an unconstrained choice by a median country with sufficiently low crisis probability relative to $\breve{\pi}(\mathcal{M})$, or a constrained choice of a higher median crisis probability country. In the previous subsection we concluded that it is π^m and not μ which drives the equilibrium Fund size. We can now qualify and strengthen this conclusion: π^m drives the equilibrium Fund size, providing π^m is low enough; otherwise the median voter will be constrained and π^1 will determine an upper limit on the size of the Fund. Once again, if the distribution of crisis probabilities becomes more skewed, so that π^1 and π^m falls relative to μ, as is consistent with the stylised facts, then we would expect the Fund to provide less coinsurance in the political equilibrium.

5.3 Qualified-majority voting and agenda setting

As noted already, in practice revisions to quotas require the support of members holding at least 85% of the votes. What implications does this have for the political equilibrium?

In the related work of Alesina *et al* (2001), the authors define qualified-majority voting (QMV) as a situation where, in a union of N members, any policy changes require a majority Q, where $1 \geqslant Q/N \geqslant 1/2$. They find that under QMV no single policy outcome unambiguously emerges from voting against all alternatives. In the context of our model, if g_{N-Q} and g_Q are the Fund sizes preferred by the $N-Q$ and Q ranked countries (in increasing order of π) then the set of options $\{g_{N-Q}; \ldots; g_Q\}$ cannot be beaten under QMV by an alternative option.[27] The size of the potential winning set is an increasing function of

low enough that the cut-off country would not hold any reserves in the no-Fund world then $\frac{\partial \breve{\pi}(g^u(\pi^i))}{\partial g^u} > 0$ holds for any general utility function.

[27] In the case of the Fund, which requires a 85% qualified majority, with the current 185 members we have $Q = 158$ and $N - Q = 27$. So the optimal choices of Fund size chosen by countries (ranked by increasing crisis probability) 1 through to 27 and from 159 through to 185 can all be defeated by an alternative with an 85% qualified majority. The choices of countries 28 through to 158 cannot. For example, g_{28} cannot be beaten by g_{29} by the required 85% majority, as all countries from 29 to 185 would prefer g_{29}, providing a majority of only 84.9%.

the required majority Q.[28]

However, Alesina *et al* find that in this type of situation the ambiguity is resolved by an agenda setter who decides which alternatives are put to a vote. In the context of our model, if there is only one round of proposals, with no amendments allowed, then the agenda setter will make a proposal which maximises her expected utility subject to the incentive constraint that at least $Q-1$ other countries weakly prefer the new Fund size to the *status quo*.

As the United States has a veto power on QMV decisions at the Fund (with more than $Q-N$ votes) it can block any proposal by an agenda setter with which it disagrees. This would seem to increase the *status quo* bias against any enlargement of the Fund. However, in the Fund's case the agenda setter is perhaps most likely to be the United States itself, given that it is the largest shareholder and perhaps also because the Fund is based in the United States, which may increase the political influence the United States can exert over the IMF's staff and its Executive Board. Suppose the United States is the sole agenda setter and that the United States is also the member with the lowest vulnerability to a crisis. In this scenario the initial size of the Fund will be chosen by the United States to maximise its expected utility. This will be preferred by the rest of the membership, compared with the option of no Fund. As the agenda setter the United States can prevent any other options from being put to the vote. Over time, assuming that the United States remains both the agenda setter and the member with the lowest crisis probability, it will be able to increase the Fund size, should it want to do so, as this will gain the consent of a sufficiently large majority of the membership, but it will not be able to reduce it, in the absence of a generalised reduction of crisis probabilities.

Thus, the conclusion we reach again strengthens those of the previous subsections: if the member with the lowest crisis probability is the agenda setter, it is π^1 that drives the equilibrium Fund size. If the distribution of crisis probabilities becomes more skewed, so that π^1 falls relative to μ, we would expect the equilibrium Fund subscription level to fall.

In Appendix A we consider how the existence of spillovers might modify these conclusions. In particular, we show how spillovers can lead countries to prefer a larger Fund size, other things being equal. However, this should not detract from the key conclusion reached from the analysis of this section, which is essentially that the Fund size is likely to be driven by countries with crisis probabilities that are, perhaps considerably, below the mean for the membership as a whole. What it does demonstrate, however, is that even if these countries themselves have very low, or possibly zero crisis probabilities, the existence of spillovers can rationalise a revealed preference for a positive Fund size, even if this is still considerably below that which might be preferred by, for example, the member with a mean crisis probability.

[28] Assume that the alternative option is the maximum public good provision unanimously supported against no provision: then, the lower bound of the winning set is decreasing in the required majority moving from the simple majority voting level of g_m to the unanimity level of g_0. The upper bound is increasing in the required majority for Q small enough and decreasing for Q big enough.

6 Welfare analysis

In this section we consider the conditions under which there is scope for Pareto-improving changes in the Fund size and the associated subscriptions of members. The intention is to illustrate why the framework underpinning the IMF's lending operations might need reconsidering, rather than to advocate a particular new approach.

Consider a given interior political equilibrium choice of Fund subscription, g, with corresponding Fund size and crisis payout, f. Can we change the Fund size and *individual* subscriptions to make at least one country better off and no country worse off? Note that by framing the question in this way we introduce the possibility that members pay different subscriptions, although we assume they still receive equal payouts in the event of a crisis.

Revenue neutrality requires that $\Delta f = \frac{1}{\mu N}\sum_{j=1}^{N}\Delta g^j$. Denote the modified subscription level of country i as $g^{i'} = g + \Delta g^i$ with the modified Fund payout common across countries as $f' = f + \Delta f = g/\mu + \Delta f$. We first consider how the consumption of country i is affected in crisis and non-crisis states by Δg^i and Δf, before considering what this implies for that country's expected utility.

We can write the following general expressions for consumption in each state for country i, in terms of g, $f = g/\mu$, Δg^i and Δf:

$$\begin{aligned} c_c^{i,f}(g^{i'}, f', \pi^i) = (1-\delta)\,\rho_A\left[1 - b^{i,f*} - \left(g + \Delta g^i\right)\right] + \rho_C b^{i,f*} \\ + (\rho_C - \rho_F)\,(g/\mu + \Delta f) + \rho_F\left(g + \Delta g^i\right) \end{aligned} \tag{5}$$

$$c_n^{i,f}(g^{i'}, f', \pi^i) = \rho_A\left[1 - b^{i,f*} - \left(g + \Delta g^i\right)\right] + \rho_R b^{i,f*} + \rho_F\left(g + \Delta g^i\right) \tag{6}$$

where $b^{i,f*}$ is a non-increasing function of the new Fund subscription of country i. It follows that for all values of π^i the following conditions hold:

$$\begin{aligned} &\tfrac{\partial c_c^{i,f}}{\partial \Delta g^i} = \tfrac{\partial c_c^{i,f}}{\partial g} - \tfrac{(\rho_C - \rho_F)}{\mu} \qquad && \tfrac{\partial c_n^{i,f}}{\partial \Delta g^i} = \tfrac{\partial c_n^{i,f}}{\partial g} \\ &\tfrac{\partial c_c^{i,f}}{\partial \Delta f} = (\rho_C - \rho_F) && \tfrac{\partial c_n^{i,f}}{\partial \Delta f} = 0 \end{aligned}$$

Expected utility is defined as $W^i = \pi^i u(c_c^{i,f}) + (1-\pi^i)u(c_n^{i,f})$. Given the partial derivatives of consumption in each of the states, for all values of π^i we have:

$$\frac{\partial W^i}{\partial \Delta g^i} = \frac{\partial W^i}{\partial g} - \frac{(\rho_C - \rho_F)}{\mu}\pi^i u'\left(c_c^{i,f}\right)$$

$$\frac{\partial W^i}{\partial \Delta f} = (\rho_C - \rho_F)\,\pi^i u'\left(c_c^{i,f}\right)$$

For country i, the change in welfare is equal to $\Delta W^i = W^i(g^{i'}, f', \pi^i) - W^i(g, f, \pi^i)$. For small changes in payouts and subscription levels this can be approximated by:

$$\Delta W^i = \Delta g^i \frac{\partial W^i}{\partial \Delta g^i} + \Delta f \frac{\partial W^i}{\partial \Delta f} \tag{7}$$

Putting the above expressions into equation 7 yields:

$$\Delta W^i = \Delta g^i \frac{\partial W^i}{\partial g} + \left(\Delta f - \frac{\Delta g^i}{\mu} \right) (\rho_C - \rho_F) \pi^i u' \left(c_c^{i,f} \right) \tag{8}$$

This expression is useful to characterise the likely sign of ΔW^i since we know the range of π^i over which $\partial W^i / \partial g$ is positive or negative. From the Proof of Proposition 4 we know that $\partial W^i / \partial g$ is increasing in π^i. Denote by h the member for which $\partial W^h / \partial g = 0$. With unconstrained majority voting $h = m$, but with either constrained majority voting, or QMV with an agenda setter, as described in the previous section, then $h < m$. It follows that for $i < h$ then $\partial W^i / \partial g < 0$ and for $i > h$ then $\partial W^i / \partial g > 0$.

From equation 8 we can make the following inferences. First, all members would prefer to pay a higher subscription to bring about a (small) increase in the Fund size, providing the increase in their own subscription is not too high. This follows as the coefficient on Δf is necessarily positive, given $u'(c_c^{i,f}) > 0$, and so ΔW^i will be positive providing Δg^i is not too large. Conversely, all members would prefer a lower Fund size, providing their subscription falls sufficiently.

Second, starting from any given political equilibrium, there is no common increase or decrease in the Fund subscription that is Pareto improving. A common change in subscription requires that $\Delta g^i = \mu \Delta f$ for all i and so the second term in 8 is zero. Following an increase (decrease) in Fund size the first term is negative (positive) for any $i < h$, but positive (negative) for $i > h$, and so benefits high-risk (low-risk) members at the expense low-risk (high-risk) members.

Third, for member h it must be the case that $\Delta W^h > 0$, providing $\Delta g^h < \mu \Delta f$, that is, providing any increase (decrease) in subscription for member h is less (greater) than the average for the Fund membership as a whole.

Now suppose we restrict the scheme used to finance a change in the Fund size to be linear in the crisis probability, such that $\Delta g^i = k\pi^i$, where k is necessarily a monotonically increasing function of Δf.[29] Then we can write:

$$\Delta W^i = k\pi^i \left\{ \frac{\partial W^i}{\partial g} + \frac{1}{\mu} \left(\mu - \pi^i \right) (\rho_C - \rho_F) u' \left(c_c^{i,f} \right) \right\} \tag{9}$$

If $\pi^h < \mu$, as we would expect under each of the political equilibria outlined in the previous section and which we assume in the remainder of this section, then it follows that a small increase in Fund size raises the welfare of all members with crisis probability $\pi^h < \pi^i < \mu$.[30] Conversely, a decrease in the Fund size reduces the welfare of these same members, and so is *never* Pareto improving.

Note that for $\pi^i < \pi^h$ an increase in the Fund size still improves the welfare of member i, providing π^i is sufficiently close to π^h. This is because in this case the negative term in 9 is of second-order magnitude, whereas the positive term is of first-order magnitude. Similarly, for $\pi^i > \mu$ an increase in the Fund size still improves the welfare of member i, providing π^i is sufficiently close to μ.

[29] If the increase in the Fund size is to be adequately financed we require that $k \sum_i \pi^i = \Delta f$.

[30] Similarly, in the unlikely case where $\pi^h > \mu$ a decrease (increase) in Fund size raises (reduces) the welfare of all members with crisis probability $\pi^h > \pi^i > \mu$.

By imposing some further restrictions we can show that a small increase in Fund size benefits all members for sure.

Proposition 6 *If $\Delta g^i = k\pi^i$ then (a) $\rho_C \geq \rho_A$ is a sufficient, but not a necessary condition for a small increase in Fund size to improve the welfare of all members with $\pi^i < \pi^h$ and (b) $\rho_F \geq \rho_R$ is a sufficient, but not a necessary condition for a small increase in Fund size to improve the welfare of all members with $\pi^i > \mu$.*

Proof. See Appendix B.6 ■

Taken together, these results mean that under these assumptions small increases in the Fund size, financed by the rule $\Delta g^i = k\pi^i$, are necessarily Pareto improving. The key to this result is that the linear financing rule distributes the cost of the increase in Fund size so that those that benefit most from it – that is, those with a high crisis probability – pay proportionately more than those who benefit less.

7 Conclusions

This paper develops a simple one-period investment model in which countries can protect themselves against the risk of adverse shocks by subscribing to a credit union or by accumulating reserves. The financial structure of the credit union mimics that of the IMF, crucially in that its overall size is determined by a vote of the membership. We assume that countries are equal in all respects except in their vulnerability to a crisis. This allows us to isolate the impact of the increasing heterogeneity of the Fund's membership, in terms of vulnerability to a crisis, on its effectiveness as a provider of consumption smoothing over crisis states. Adding other aspects of country heterogeneity, eg size and returns, is clearly an important avenue to pursue in subsequent work.

Our simple model yields some useful insights. If we accept that IMF member countries in 1946 were broadly similar, our analysis suggests that the Fund's founding fathers created an institution that was fit for purpose. Moreover by giving members the opportunity to revisit the size of the Fund every five years, they created a mechanism to ensure that the size of the Fund could be modified so that it continued to provide the appropriate amount of crisis-state payouts for a homogeneous, but crisis-prone membership.

However, based on our model, this adjustment mechanism may no longer work so well. Nowadays the Fund's membership consists of creditor and would-be debtor groups. In our model the equilibrium choice of the size of the Fund is likely to be driven by the preferences of creditor countries with a relatively low crisis probability. We make essentially this same inference from each of the political equilibria identified in Section 5. Under unconstrained majority voting it is the median voter that is decisive and over the life of the Fund it is likely that the median crisis probability has fallen relative to the mean among its membership. If the median voter is constrained by a binding participation constraint then this will further limit the size of the Fund. Finally, if the Fund size is determined by an agenda setter with a low crisis probability, this also limits the size of the Fund in our model.

This result has several implications, each of which we can observe in the global economy, although clearly there are other potential explanations. First, high crisis probability countries are likely to increasingly turn to self-insurance and hold more reserves than before. Second, in those regions where second-round spillovers are larger than average, regional Funds are likely to develop to provide a 'second-line' of multi-lateral crisis insurance. Both of these features have been observed in Asia. And finally, the average size of Fund assistance to actual crisis countries is likely to increase as the proportion of countries at risk of crisis falls. This too has been observed, and to some extent should offset the incentive for risky countries to self-insure.

These results could be taken as implying the Fund should be increased in size. We would caution against rushing to this conclusion. For a start, without changing the structure of the Fund, such a conclusion risks wishing away the problem, which is rooted in the institutional constraints which limit the size of the Fund. But more fundamentally, our model does not take into account moral hazard – we have made no allowance in our model for a relationship between the crisis probability and the size of Fund assistance. This is potentially an important omission: as the debates of recent years have demonstrated, many commentators have been deeply concerned about the risk of moral hazard associated with large Fund financial programmes.

We draw a different conclusion: that the framework governing the Fund's lending operations may no longer be appropriate. The credit union model that underpins the Fund's structure made sense in 1946 when the Fund was comprised of similar countries. That may no longer be the case. An alternative approach may be needed: one which takes into account that creditor and debtor countries have different interests and which takes into account the moral hazard consequences of large-scale lending. The ongoing international debate about the strategic direction of the IMF could helpfully encompass this issue.

A Appendix: Spillovers

Cross-country crisis spillovers can be represented in a reduced-form by each country caring about the consumption of others. This formulation can pick up economic or geopolitical reasons why countries may care about the consumption levels of others. This leads to a modified expected utility function for country i of:

$$W^i = \pi^i u(c_c^{i,f}) + (1-\pi^i)u(c_n^{i,f}) + \sum_{j=1,j\neq i}^{N} \beta_j^i \left(\pi^j v(c_c^{j,f}) + (1-\pi^j)v(c_n^{j,f})\right) \quad (10)$$

where country i cares about the expected consumption of another country j through the function $v(\cdot)$ with a weight β_j^i. This allows spillovers to be specific to country pairs. Note that with this formulation, *reserve* choices are not influenced by spillovers, as the reserves held by country i have no impact on the consumption of country j. Thus Propositions 1 and 2 are unaffected. However, the size of the Fund does impact on the consumption of all member countries. Consequently, preferences over Fund size are affected by the introduction of spillovers.

Consider Proposition 3 concerning a country's preference over the Fund size. What impact do spillovers have on individual country preferences over g? In order to proceed, we make two simplifying assumptions. First, assume that country i cares about country j's consumption in the same way that country j does (ie $v(\cdot)$ and $u(\cdot)$ are the same function). Second, let country i care about country j's consumption with a weight $\beta_j^i = \beta^i l(\pi^j) \leq 1$ for all i,j. Define $Z^i(g) = \pi^i u(c_c^{i,f}) + (1-\pi^i)u(c_n^{i,f})$. This is the expected utility country i receives from its own consumption alone. Note that in the absence of spillovers $W^i = Z^i(g)$. The new Lagrangean for the preferred choice of g^i is:

$$\max_{g^i} \mathcal{L} = Z^i(g^i) + \sum_{j=1,j\neq i}^{N} \beta^i l(\pi^j) Z^j(g^i) + \lambda_1 b^{i,f*}(g^i) + \lambda_2 g^i - \lambda_3 (b^{i,f*}(g^i) + g^i - 1)$$

The first-order condition for the interior solution, g^{i*}, is:[31]

$$0 = Z^{i\prime}(g^i) + \beta^i \sum_{j=1,j\neq i}^{N} l(\pi^j) Z^{j\prime}(g^i) \quad (11)$$

In Subsection 4.2 we showed that, in the model without spillovers, $g^{i*} \geq 0$ and $b^{i,f*}(g^{i*}) = 0$ providing $\mu < \tilde{\mu}$. In the model with spillovers a similar condition on the level of the mean crisis probability can be derived. Moreover, by using the same method as in the no-spillovers case we can also show that the single-crossing condition holds and hence we can apply the median voter theorem.

[31] The second-order conditions are satisfied under weak assumptions on the concavity of the utility function.

The key question is how the magnitude of the spillovers, captured by the summary statistic β^i, impacts on the optimum choice of Fund subscription levels. We know from the proof of Proposition 4 that, holding g fixed, $Z^j(g)$ increases with j, and so if g^i is such that $Z^{i\prime}(g^i) = 0$ then $Z^{j\prime}(g^i) < 0$ for $j < i$ and $Z^{j\prime}(g^i) > 0$ for $j > i$. Consequently, we can reach two conclusions. First, if country i cares about consumption of all other countries equally, so that $l(\pi^j) = 1$ for all j, then this will bunch together country preferences over g and this bunching will be more pronounced as β^i increases for all i. To see this, consider the extreme case where $\beta^i = 1$ for all i, which means that all countries care about each others' consumption as much as they care about their own. In this situation the first-order condition 11 is identical for all Fund members and so accordingly is the preferred Fund size. We can therefore conclude that stronger spillovers are, other things being equal, likely to raise the political equilibrium Fund size, where this is driven by a member with a below-mean crisis probability, such as is the case in each of the political equilibria outlined in Section 5. Second, if countries only care about the consumption of other countries which have a *higher* crisis probability, so that $l(\pi^j) = 0$ for $j < i$ and $l(\pi^j) > 0$ for $j > i$, then stronger spillovers unambiguously raise the Fund size preferred by all countries. Under this assumption stronger spillovers will have an unambiguously positive impact on the political equilibrium Fund size.

B Appendix: Proofs

B.1 Proposition 1 – Choice of reserves in world with no Fund

From the first-order condition for the maximisation of equation 2 we obtain an implicit expression for the interior solution for the optimal reserve holdings for country i (satisfying the second-order condition):

$$\pi^i \Gamma u'(\rho_A(1-\delta)(1-b^{i,nf*})+\rho_C b^{i,nf*}) = (1-\pi^i)(\rho_A-\rho_R)u'(\rho_R b^{i,nf*}+\rho_A(1-b^{i,nf*}))$$

where $\Gamma \equiv \rho_C - (1-\delta)\rho_A$. Using the implicit function theorem we can see that the partial derivative of $b^{i,nf*}$ with respect to π^i is strictly increasing for concave utility functions.

The first-order condition implies corner solutions with $b^{i,nf*} = 0$ for $\pi^i \in [\pi^1, \underline{\pi}^{nf})$ and $b^{i,nf*} = 1$ for $\pi^i \in (\bar{\pi}^{nf}, \pi^N]$ where:

$$0 < \underline{\pi}^{b,nf} \equiv [1 + \frac{\Gamma u'((1-\delta)\rho_A)}{(\rho_A - \rho_R)u'(\rho_A)}]^{-1} < \bar{\pi}^{b,nf}$$

$$\bar{\pi}^{b,nf} \equiv [1 + \frac{\Gamma u'(\rho_C)}{(\rho_A - \rho_R)u'(\rho_R)}]^{-1} < 1$$

B.2 Proposition 2 – Choice of reserves for given Fund size g

The non-median countries face the constrained optimisation problem expressed as the Lagrangean of equation 3. The first-order condition with respect to $b^{i,f}$, given g, is as follows (with the second-order condition satisfied):

$$\pi^i \Gamma u'((1-\delta)\rho_A + \Gamma b^{i,f} + \Omega g)-$$
$$(1-\pi^i)(\rho_A - \rho_R)u'(\rho_A - (\rho_A - \rho_R)b^{i,f} - (\rho_A - \rho_F)g) + \lambda_1 - \lambda_3 = 0 \qquad (12)$$

where λ_1 and λ_3 are the Lagrange multipliers on the constraints $b^{i,f} \geq 0$ and $b^{i,f} + g^* \leq 1$ respectively. The interior solution for $b^{i,f*}$ is implicitly defined by 12 with $\lambda_1 = \lambda_3 = 0$. With a concave utility function, for the interior solution $\partial b^{i,f*}/\partial \pi^i > 0$ and $\partial b^{i,f*}/\partial g^* < 0$.

Turning to the corner solution with $b^{i,f} = 0$, the requirement that $\lambda_1 > 0$ implies the condition:

$$\pi^i < \underline{\pi}^{b,f} \equiv [1 + \frac{\Gamma u'((1-\delta)\rho_A + \Omega g^*)}{(\rho_A - \rho_F)u'(\rho_A - (\rho_A - \rho_F)g^*)}]^{-1}$$

If the median country holds no reserves then since $\partial b^{i,f*}/\partial \pi^i > 0$ it can be seen that $\pi^m < \underline{\pi}^{b,f}$.

For the other corner solution, $b^{i,f*} + g^* = 1$, $\lambda_3 > 0$ implies the condition:

$$\pi^i > \bar{\pi}^{b,f} \equiv [1 + \frac{\Gamma u'(\rho_C + g^*(\rho_C - \rho_F)(1/\mu - 1))}{(\rho_A - \rho_F)u'(\rho_R - g^*(\rho_R - \rho_F))}]^{-1}$$

B.3 Proposition 3 – Country choice of reserves and Fund size

First let us show the equivalence of conditions on mean crisis probability and on relative returns.

Using the definitions of Ω and Γ:

$$\frac{\Omega}{(\rho_A - \rho_F)} > \frac{\Gamma}{(\rho_A - \rho_R)} \iff \Gamma(\rho_A - \rho_R) + (\rho_C - \rho_F)(\rho_A - \rho_R)(1/\mu - 1) > \Gamma(\rho_A - \rho_F)$$

$$\iff \mu < \tilde{\mu} \quad \text{where} \quad \tilde{\mu} \equiv \{1 + \frac{(\rho_C - (1-\delta)\rho_A)(\rho_R - \rho_F)}{(\rho_C - \rho_F)(\rho_A - \rho_R)}\}^{-1}$$

Similarly $\frac{\Omega}{(\rho_A - \rho_F)} < \frac{\Gamma}{(\rho_A - \rho_R)} \iff \mu > \tilde{\mu}$ and $\frac{\Omega}{(\rho_A - \rho_F)} = \frac{\Gamma}{(\rho_A - \rho_R)} \iff \mu = \tilde{\mu}$

The first-order conditions of the optimisation problem represented in the Lagrangean of 4 with respect to the choice of $g^{i,f}$ is as follows (with second-order conditions satisfied and associated complementary slackness conditions applying):

$$\pi^i \Omega u'(c_c^{i,f}) - (1-\pi^i)(\rho_A - \rho_F)u'(c_n^{i,f}) + \lambda_2 - \lambda_3$$
$$+\frac{\partial b^{i,f*}}{\partial g^{i,f}}(\pi^i \Gamma u'(c_c^{i,f}) - (1-\pi^i)(\rho_A - \rho_R)u'(c_n^{i,f}) + \lambda_1 - \lambda_3) = 0 \qquad (13)$$

However, substituting in from the first-order conditions for the optimal choice of reserves, given Fund size, we obtain the following condition.

$$\pi^i \Omega u'(c_c^{i,f}) - (1-\pi^i)(\rho_A - \rho_F)u'(c_n^{i,f}) + \lambda_2 - \lambda_3 = 0 \tag{14}$$

The optimality conditions are thus specified by equations 14 and 12.

Case 1: $\mu < \tilde{\mu} \Longleftrightarrow \frac{\Omega}{(\rho_A - \rho_F)} > \frac{\Gamma}{(\rho_A - \rho_R)}$

Rearranging the first-order conditions for the choice of Fund subscription levels and reserves, we obtain:

$$\frac{\pi^i \Omega u'(c_c^{i,f})}{(1-\pi^i)(\rho_A - \rho_F)u'(c_n^{i,f})} = 1 + \frac{\lambda_3 - \lambda_2}{(1-\pi^i)(\rho_A - \rho_F)u'(c_n^{i,f})}$$

$$\frac{\pi^i \Gamma u'(c_c^{i,f})}{(1-\pi^i)(\rho_A - \rho_R)u'(c_n^{i,f})} = 1 + \frac{\lambda_3 - \lambda_1}{(1-\pi^i)(\rho_A - \rho_R)u'(c_n^{i,f})}$$

Given $\mu < \tilde{\mu}$ and the above two expressions we obtain the condition that

$$\frac{\lambda_3 - \lambda_2}{(\rho_A - \rho_F)} > \frac{\lambda_3 - \lambda_1}{(\rho_A - \rho_R)}$$

Consider first interior solutions with positive investment (ie $b^{i,f} + g^{i,f*} < 1$, $\lambda_3 = 0$). The above inequality simplifies to $\frac{\lambda_1}{(\rho_A - \rho_R)} > \frac{\lambda_2}{(\rho_A - \rho_F)}$. If we have positive Fund choice ($\lambda_2 = 0$) then the choice of reserves is zero ($\lambda_1 > 0$). The interior solution for $g^{i,f*}$ is implicitly defined by the following expression:

$$\pi^i \Omega u'((1-\delta)\rho_A + \Omega g^{i,f*}) = (1-\pi^i)(\rho_A - \rho_F)u'(\rho_A - (\rho_A - \rho_F)g^{i,f*}) \tag{15}$$

If the Fund choice is zero then the reserve choice must also be zero. The reserve choice cannot be positive – if it was then λ_2 would have to be negative which cannot be the case as the Lagrange multiplier is greater than or equal to zero. Thus the upper corner solution in this case must be that $g^* = 1$ and reserves are zero. We now consider the two corner solutions.

Consider the corner solution $g^{i,f*} = 0$. From the requirement $\lambda_2 > 0$, the condition for countries to choose zero Fund size in this case is:

$$\pi^i < \underline{\pi}^{g,f} \equiv [1 + \frac{\Omega u'((1-\delta)\rho_A)}{(\rho_A - \rho_F)u'(\rho_A)}]^{-1}$$

Now, consider the corner solution $g^{i,f*} = 1$. By similar analysis we obtain the condition that the crisis probability be high enough such that:

$$\pi^i > \bar{\pi}^{g,f} \equiv [1 + \frac{\Omega u'(\rho_F + ((\rho_C - \rho_F))/\mu)}{(\rho_A - \rho_F)u'(\rho_F)}]^{-1}$$

We thus have $0 < \underline{\pi}^{g,f} < \bar{\pi}^{g,f} < 1$.

Case 2: $\mu > \tilde{\mu} \Longleftrightarrow \frac{\Omega}{(\rho_A - \rho_F)} < \frac{\Gamma}{(\rho_A - \rho_R)}$

We follow a similar approach to Case 1. Using the first-order condition and the reserve relations we obtain:

$$\frac{\lambda_3 - \lambda_2}{(\rho_A - \rho_F)} < \frac{\lambda_3 - \lambda_1}{(\rho_A - \rho_R)}$$

Consider first interior solutions with positive investment (ie $b^{i,f} + g^{i,f*} < 1$, $\lambda_3 = 0$). The above inequality simplifies to $\frac{\lambda_1}{(\rho_A - \rho_R)} < \frac{\lambda_2}{(\rho_A - \rho_F)}$. A positive Fund choice would imply a negative Lagrange multiplier $\lambda_1 = 0$ and so this case cannot apply. If the choice of reserves is positive then the optimal Fund subscription level is zero. The binding cut-offs for the corner solutions of zero reserves and full reserve holdings are as in Section B.1. If reserves are zero then the optimal choice of Fund size is also zero.

Case 3: $\mu = \tilde{\mu} \Longleftrightarrow \frac{\Omega}{(\rho_A - \rho_F)} = \frac{\Gamma}{(\rho_A - \rho_R)}$

Substituting the reserve relation into 14 and 12 and simplifying we obtain:

$$\begin{aligned}\pi^i(\Omega - \Gamma)u'((1-\delta)\rho_A + \Gamma b^{i,f*} + \Omega g^{i*}) \\ = (1-\pi^i)(\rho_R - \rho_F)u'(\rho_A - (\rho_A - \rho_R)b^{i,f*} - (\rho_A - \rho_F)g^{i*}) \quad (16)\end{aligned}$$

This relation can only hold if $\rho_R > \rho_F$. Furthermore the reserve relation implies $\lambda_1 = \lambda_2$. The only consistent possibilities are that both $b^{i,f*}$ and $g^{i,f*}$ are zero or both are positive. Both are zero if $\pi^i < \left(1 + \frac{\Omega u'((1-\delta)\rho_A)}{(\rho_A - \rho_F)u'(\rho_A)}\right)^{-1}$. The other corner solution of $b^{i,f*} + g^{i,f*} = 1$ cannot hold.[32] Thus, if both are positive they are under-determined.

B.4 Proposition 4 – Political equilibrium choice over Fund size

Denote country i's expected utility from a Fund size of g as $W(g, \pi^i)$. Following Ashworth and Bueno de Mesquita (2006), 'increasing differences' in the return from changing policy is a sufficient condition for single-crossing. 'Increasing differences' holds if the return from changing policy is increasing in country type, ie

$$W(g, \pi^i) - W(g', \pi^i) \geq W(g, \pi^{i'}) - W(g', \pi^{i'})$$

$\forall g > g'$ and $\forall \pi^i > \pi^{i'}$. For small changes in π^i, this condition can be approximated by:

$$(\pi^i - \pi^{i'})[\frac{\partial W(g, \pi^i)}{\partial g} - \frac{\partial W(g, \pi^{i'})}{\partial g}] \geq 0$$

[32] This can be seen from the condition for $\lambda_3 > 0$ obtained from 14. Substituting in from 16 yields an inconsistent condition.

This is equivalent to $\frac{\partial^2 W(g,\pi^i)}{\partial g \partial \pi^i} \geq 0$. With $\frac{\partial W(g,\pi^i)}{\partial g} = \pi^i \Omega u'(c_c^f) - (1-\pi^i)(\rho_A - \rho_F)u'(c_n^f)$ we have,

$$\begin{aligned}\frac{\partial^2 W(g,\pi^i)}{\partial g \partial \pi^i} &= \Omega u'(c_c^f) + (\rho_A - \rho_F)u'(c_n^f) - \\ &\quad \frac{\partial b^{i,f}}{\partial \pi^i}[\pi^i \Gamma \Omega u''(c_c^f) + (1-\pi^i)(\rho_A - \rho_R)(\rho_A - \rho_F)u''(c_n^f)]\end{aligned}$$

Thus $\frac{\partial^2 W(g,\pi^i)}{\partial g \partial \pi^i} \geq 0 \ \forall \pi^i > \pi^{i'}$. We therefore have strictly increasing differences. This is a sufficient condition for single-crossing and hence for the application of the median voter theorem.

B.5 Proposition 5 – Utility comparison in and outside the Fund

From the proof of Proposition 4, we have

$$W(g,\pi^i) - W(g',\pi^i) > W(g,\pi^{i'}) - W(g',\pi^{i'})$$

$\forall g > g'$ and $\forall \pi^i > \pi^{i'}$. Setting $g' = 0$ this equation tells us that the gain in welfare from being in the Fund compared to being outside is strictly increasing in a country's crisis probability.

Consider the extreme case where $\pi_1 = 0$. Such a country would hold zero reserves in both the Fund and no-Fund worlds. Expected utility of this country is just its utility in the non-crisis state and so $W(g,\pi_1 = 0) - W(0,\pi_1 = 0) = u(c_n^{1,f}) - u(c_n^{1,nf})$. In the no-Fund world non-crisis state consumption is $c_n^{1,f} = \rho_A$. Non-crisis state consumption in the Fund world is lower due to the required subscription to the Fund $c_n^{1,nf} = \rho_A - (\rho_A - \rho_F)g^*$. Thus the country is better off outside the Fund. The median country is better off in the non-zero Fund (since it had the option of choosing a zero Fund size). Given that the welfare difference between being in and outside the Fund is continuous and strictly increasing in the crisis probability we have a unique fixed point $\widehat{\pi}$ at which $W(g,\widehat{\pi}) = W(0,\widehat{\pi})$. So, provided that $\pi_1 \leq \widehat{\pi}$ we have at least one country who would be better off outside the Fund.

B.6 Proposition 6 – Pareto-improving increase in Fund size with a linear financing scheme

Partial differentiation of 5 and 6 gives:

$$\frac{\partial c_c^{i,f}}{\partial g} = \Gamma\left(1 + \frac{\partial b^{i,f*}}{\partial g}\right) + \left(\frac{1-\mu}{\mu}\right)(\rho_C - \rho_F) \tag{17}$$

$$\frac{\partial c_n^{i,f}}{\partial g} = \rho_F - \rho_R - (\rho_A - \rho_R)\left(1 + \frac{\partial b^{i,f*}}{\partial g}\right) \tag{18}$$

Using 17, 18 and 9 the impact on expected utility of a small increase in Fund size can be written as:

$$\begin{aligned}\Delta W^i = k\pi^i \Big\{ & \left(1+\tfrac{\partial b^{i,f*}}{\partial g}\right)\left[\pi^i \Gamma u'\left(c_c^{i,f}\right) - (1-\pi^i)(\rho_A-\rho_R)u'\left(c_n^{i,f}\right)\right] \\ & + (1-\pi^i)\left[(\rho_C-\rho_F)\,u'\left(c_c^{i,f}\right) + (\rho_F-\rho_R)\,u'\left(c_n^{i,f}\right)\right]\Big\}\end{aligned} \tag{19}$$

Consider first part (a) of the proposition. As $\pi^h < \underline{\pi}^{b,f}$ we know from Proposition 2 that $\partial b^{i,f*}/\partial g = 0$. Consequently 19 becomes:

$$\Delta W^i = k\pi^i \left\{ \left[\pi^i\Gamma + (1-\pi^i)\,(\rho_C-\rho_F)\right] u'\left(c_c^{i,f}\right) - (1-\pi^i)\,(\rho_A-\rho_F)\,u'\left(c_n^{i,f}\right)\right\}$$

Given $c_n^{i,f} > c_c^{i,f}$ implies $u'\left(c_c^{i,f}\right) > u'\left(c_n^{i,f}\right)$, a sufficient, but not a necessary condition for $\Delta W^i > 0$ is that

$$\pi^i\Gamma + (1-\pi^i)\,(\rho_C-\rho_A) \geq 0$$

Given $\Gamma > 0$ a sufficient, but not a necessary condition for this to hold is that $\rho_C \geq \rho_A$, which proves part (a) of the proposition.

Now consider part (b) of the proposition. For $\pi^i > \bar{\pi}^{b,f}$ we know that $b^{i,f*} + g = 1$ and so $\partial b^{i,f*}/\partial g = -1$. In this case a positive marginal utility from consumption in each of the states is sufficient to ensure that $\Delta W^i > 0$. For $\mu < \pi^i < \bar{\pi}^{b,f}$ then from 12, using the implicit function theorem:

$$1 + \frac{\partial b^{i,f*}}{\partial g} = -\frac{N}{D}$$

where

$$N = -\left[\pi^i\Gamma\,(\rho_C-\rho_F)\left(\frac{1-\mu}{\mu}\right)u''\left(c_c^{i,f}\right) - (1-\pi^i)(\rho_A-\rho_R)\,(\rho_F-\rho_R)\,u''\left(c_n^{i,f}\right)\right]$$

$$D = -\left[\pi^i\Gamma^2 u''\left(c_c^{i,f}\right) + (1-\pi^i)(\rho_A-\rho_R)^2 u''\left(c_n^{i,f}\right)\right] > 0$$

By substitution into 19 this becomes:

$$\begin{aligned}\Delta W^i = \tfrac{k\pi^i}{D} \Big\{ & \left[(1-\pi^i)D\,(\rho_C-\rho_F) - N\pi^i\Gamma\right] u'\left(c_c^{i,f}\right) \\ & + (1-\pi^i)\left[N(\rho_A-\rho_R) + D\,(\rho_F-\rho_R)\right] u'\left(c_n^{i,f}\right)\Big\}\end{aligned} \tag{20}$$

The concavity of the utility function means that $\rho_F \geq \rho_R$ is a sufficient, but not a necessary condition to ensure that each of the terms in square brackets in 20 is positive. This implies $\Delta W^i > 0$ for all $\pi^i > \mu$ and thus completes the proof.

References

[1] **Alesina, A, Angeloni, I and Etro, F (2001),** 'Institutional rules for federations', *NBER Working Paper No. 8646*.

[2] **Alesina, A, Angeloni, I and Etro, F (2005),** 'International unions', *American Economic Review*, Vol. 95(3), pages 602-15.

[3] **Alesina, A and Perotti, R (1998),** 'Economic and political risk in fiscal unions', *The Economic Journal*, Vol. 108(449), pages 989-1,008.

[4] **Ashworth, S and Bueno de Mesquita, E (2006),** 'Monotone comparative statics in models of politics: a method for simplifying analysis and enhancing empirical content', *American Journal of Political Science*, Vol. 50(1), pages 214-31.

[5] **Chami, R, Sharma, S and Shim, I (2004),** 'A model of the IMF as a coinsurance arrangement', *IMF Working Paper No. 04/219*.

[6] **Cordella, T and Levy Yeyati, E (2006),** 'A (new) country insurance facility', *International Finance*, Vol. 9(1), pages 1-36.

[7] **Crémer, J and Palfrey, T R (2000),** 'Federal mandates by popular demand', *Journal of Political Economy*, Vol. 108(5), pages 905-27.

[8] **Gans, J S and Smart, M (1996),** 'Majority voting with single-crossing preferences', *Journal of Public Economics*, Vol. 59(2), pages 219-37.

[9] **Gouveia, M (1997),** 'Majority rule and the public provision of a private good', *Public Choice*, Vol. 93(3), pages 221-44.

[10] **Persson, T and Tabellini, G (1996a),** 'Federal fiscal constitutions: risk sharing and moral hazard', *Econometrica*, Vol. 64, pages 623-46.

[11] **Persson, T and Tabellini, G (1996b),** 'Federal fiscal constitutions: risk sharing and redistribution', *Journal of Political Economy*, Vol. 104, pages 979-1,009.

[12] **Persson, T and Tabellini, G (2000),** *Political economics*, MIT Press, Cambridge, Massachusetts.

September 2007

SHOULD BANK SUPERVISORS IN DEVELOPING COUNTRIES EXERCISE MORE OR LESS FORBEARANCE?

Patrick Honohan

Institute for International Integration Studies,
Trinity College Dublin
and CEPR

Thanks to participants at the FMG Workshop: "Cycles, Contagion and Crises" London School of Economics, June 28-29, 2007, for helpful comments.

Author's email: phonohan@tcd.ie

Abstract

Although forbearance has been associated with more costly financial crises, a trigger-happy approach to closing weak banks could also precipitate an avoidable systemic collapse. In sophisticated regulatory environments, there can be net benefits from at least occasional acts of forbearance. But we argue that three key structural weaknesses in developing countries suggest that their regulators should have less forbearance discretion. This is because financial systems in developing countries tend to have worse information, less interdependence and greater agency problems.

SHOULD BANK SUPERVISORS IN DEVELOPING COUNTRIES EXERCISE MORE OR LESS FORBEARANCE?

1. Introduction and overview

1.1 Introduction

Recent credit market pressures in several advanced economies have led to bank rescues, and to relaxed criteria for liquidity loans, and have re-awakened old debates on forbearance. But even if regulators in advanced economies can successfully forbear to intervene undercapitalized institutions without unduly damaging moral hazard consequences, it is less clear what the lessons for developing countries should be.

This paper argues that greater agency and information problems, and lower structural interdependence within financial markets, in developing countries argue for less regulatory discretion.

1.2 Overview

Forbearance got a bad name during the US Savings and Loan crisis. Explicit relaxations in capital standards as well as lengthy periods of grace designed to allow undercapitalized or insolvent S&Ls to continue in operation were followed by massive abuses, involving excessive risk-taking, looting and fraud.

The reaction of numerous commentators to these relatively well-documented events was to call for the elimination of discretion and a zero-tolerance policy for capital adequacy. Steps in that direction were adopted in the US under the FDICIA legislation, though the legislation fell well short of the absolutism advocated by some.

Indeed, there is some cross-country empirical evidence suggesting that, for crises where data on the total fiscal costs are available and, where enough is known about the policy response to be able to say whether or not there was forbearance, those countries whose regulators have exercised forbearance have experienced more costly crises.

But the fragility of a stressed financial system and the imperfection of the rules defining capital and other regulatory standards imply that unthinking adherence to mechanical rules in a crisis situation could have large systemic consequences. That is why regulators in advanced economies have retained and exercised discretion.

It is one thing to argue that a sophisticated regulatory structure operating in an accountable and independent way should be able to improve on a mechanical rules-based system, by adapting enforcement policy to contingencies, thereby avoiding a regulator-induced deepening of an incipient crisis (Goodhart, 2007, makes a convincing case). But what of regulators in developing countries where these preconditions do not prevail?. Is tying the regulators' hands more or less likely to worsen the financial stability experience in such environments?

There is surprisingly little theoretical guidance as to whether these differences, between them, argue for or against less regulatory discretion. The severity of agency problems presumably call for less agent discretion. But the problems with accounting data seem to cut both ways: on the one hand they make it easier to conceal regulatory deficiencies; on the other hand accounting data might seem to be too unreliable to be used as the basis for a mechanical intervention rule. The lower degree of intermediary interdependence may reduce the risk of regulatory action precipitating damaging contagion, but evidence here is particularly thin.

Since data deficiencies are at the heart of this question, it is especially difficult to bring credible quantification to the debate. Even after bank failures crystallize, reliable data is scarce. In particular, obtaining a convincing measure of the frequency and scale of crises that were avoided through forbearance is elusive.

Given that a zero-tolerance policy cannot be robustly defended on empirical or theoretical grounds, but taking into account that contagion is less likely to be a problem, policy for developing countries should be nuanced. Enforcement need not be mechanical, but there should be a stronger presumption of enforcement in developing countries, especially where information, agency and governance problems are thought to be severe. Measures to strengthen accountability and transparency of regulatory action should be adopted, together with a tightening of capital requirements to take account of the accounting uncertainties.

1.3 Outline

This paper begins (Section 2) by reviewing the main conceptual issues identified in the theoretical literature on capital adequacy rules and enforcement. These include analysis of the impact of regulation on (i) the *ex ante* discipline effect of bankers' expectations of forbearance on their incentives to take risks; (ii) the danger of a strict closure rule resulting in what is *ex post* inefficient liquidation; (iii) contagion effects; and (iv) the effects of alternative assignments of enforcement between different public agencies. The concerns of this literature may, however, be somewhat misplaced when it comes to developing countries.

Sections 3 to 5 discuss (and present some evidence concerning) three important dimensions of the overall environment for regulation where conditions in developing countries seem far different to those in advanced economies. First, information is extremely poor: large accounting surprises are the norm rather than the exception (Section 3). Second, financial intermediaries display less interdependence (Section 4). Third, agency problems – affecting the performance of regulators as agents of the public interest – are more severe (Section 5).

Section 6 presents theoretical considerations why these three distinctive agency, information and structural features of developing country financial systems tilt the balance of advantage against forbearance. It draws on a simple model sketched in Appendix 2.

Section 7 reviews and extends the empirical literature on the contribution of forbearance to banking crisis costs, pointing to the difficulties in obtaining decisive conclusions from available data. Concluding remarks are in Section 8.

2 Conceptual issues in the literature on regulation and forbearance

From the introductory discussion it will be clear that the topic of forbearance reaches in to many aspects of regulation. This hampers a unified theoretical treatment of the subject. Indeed, to quote Freixas and Rochet (1998), banking regulation involves "diverse issues, all of them worth devoting effort to, but so heterogeneous that no model can encompass the main issues".

In all countries, banks operate under a set of standing regulations with regard to minimum capitalization. In addition there are typically limitations on lending to related parties and other rules about risk concentrations. There may be minimum liquidity ratios. Forbearance can mean waivers of any or all of these rules. Here we will focus on capital forbearance.

Confining attention to forbearance on capital adequacy regulation narrows the focus somewhat, but not by much. Despite the growing complexity and sophistication of the models employed in this part of the literature, they tend to emphasize some aspects more than others, and arguably underplay some dimensions that are important in developing country applications.

Thus, existing models emphasize:

(i) The effect of different degrees of regulatory enforcement on bank risk-taking
(ii) Socially inefficient mid-stream liquidation due to enforcement of regulations vis-à-vis a single bank
(iii) The danger that the solvency or liquidity difficulties of one bank will create significant externalities for others, manifested through various forms of contagion
(iv) Different ways of structuring regulators' incentives through assignment of powers and responsibilities.

Most models characterize the policy issue simply as whether or not the bank should be closed. In practice a graduated response, involving a range of intermediate actions or sanctions may be applied by the regulator: the bank can be obliged to take (or refrain from taking) specific actions; the frequency and intensity of on-site inspections and auditing can be stepped-up; management can be changed and boards of directors augmented. These nuances are missing from most (though not all) of the theoretical literature.

2.1 Impact of expected forbearance on bank risk-taking: the **ex ante** *problem*

Knowing that hitting a capital floor will trigger intervention, banks will adapt their behaviour.[1] On the other hand, an expectation of regulatory forbearance will tend to reduce the behavioural effect of any regulation. This is the *ex ante* problem.

If tight capital requirements reduce risk, then easing them through expected forbearance will likely increase risk. But, as is well known since the work of Kim and Santomero (1988) and Keeley and Furlong (1990), it may be a mistake to assume that a capital adequacy requirement will always reduce banking risk. After all, imposing a binding minimum capital standard on a bank is likely to alter the structure as well as the scale of the bank's risk portfolio. Although the volume of risky assets will be lower, the bank will have altered the allocation in the direction of having riskier assets (and this effect may be so strong as to result in a net increase in the risk of bank failure).

Even in this perverse case, where imposition of capital requirements increases risk, it is conceivable that expected forbearance could increase it further. For example, in the model of Rochet (1992), bankers maximize a one-period-ahead mean-variance utility of wealth function taking into account limited liability and costs of bankruptcy, subject to satisfying a *first* period capital adequacy standard. Adding a cost of breaching the minimum capital standard in the *second* period in this model would reduce the risk taken by such a bank; but expected *second period* forbearance on the capital rule would undo that effect.[2]

---

[1] For an elegant recent modeling of this, see Elizalde and Repullo (2006)

[2] This kind of two-period model continues to be explored with more realistic and flexible specifications of banking technology. For example, the banks in Kopecky and VanHoose (2006) can vary the level of loan monitoring in order to enhance the rate of return on loans and reduce the risk of violating regulatory capital minima. Introducing a minimum capital standard here will result in lower lending but may not increase monitoring, though once the standard is in place, tightening it will increase monitoring.

Returning to the more conventional world where more capital means less risk, it may still not be optimal from the point of view of ex ante risk reduction to exclude all forbearance, considering the fact that capital can be a more costly way of funding lending. This is brought out especially by models which take the risk menu faced by the bank as being continuous (rather than assuming that bankers choose between just two levels of risk or effort). Following the suggestion by Allen and Saunders (1993) that the cost of deposit insurance effectively arises because regulators choose to forbear from prompt closure decisions, So and Wei (2004) provide such a model of such regulatory behavior, and combine it with the assumption that regulatory forbearance is accompanied by intensified frequency of auditing by the regulator.[3] The impact of the resulting moral hazard on the fair insurance premium is found to be much higher than that on the bank's equity value in simulations reported by So and Wei, highlighting the way the bank can shift risk to the deposit insurer under forbearance. (Nevertheless, their simulations suggest that a small amount of forbearance may be optimal – reflecting the difference between a discrete and continuous formulation).

Where accounting information is especially deficient, as in most developing countries, the applicability of these theories can become problematic. After all, these papers all implicitly assume that banks' actual capital position and the risk to which they are exposed are both calculable with a high degree of accuracy. This is even more evident in many studies focusing on the design of regulatory capital rules under Basel II. For example, Kerkhof and Melenberg (2004) discuss the relative merits of calibrating regulatory capital according to the Value at Risk VaR (i.e., the level of loss that can be avoided over a certain period with a given – say 99 per cent – probability) or according to the expected loss conditional on a tail event. But given the current precision of credit appraisal in most developing countries, the idea that either of these could be measured with any degree of precision let alone whether one could determine the relative precision with which they were being estimated seems

[3] Thus exemplifying the potential for graduated response to emerging problems.

fanciful.[4] Instead, as discussed in Section 3 below, capital measurement is a coarse science in developing countries.

A two-period model is limited in the degree to which it can capture the intertemporal effect of forbearance. Instead, the three-period models – widely used for analysis of liquidity crises – can help. For instance the model of Dewatripont and Tirole (1994) focuses directly on the question of a rule for closing the bank in the second period that induces the bank's managers to make adequate effort in the first period. This rule should not simply take account of the best estimate at the second period of the bank's future profit in the third period; in addition, since first period profits of the bank are indicative of the managers' effort, they should also be taken into account in the closure decision. Drawing on parallels with the modern theory of corporate finance, Dewatripont and Tirole show that the best *ex ante* rule can be implemented by giving the regulator (as representative of the depositors) control (i.e. the decision on whether or not to close) when first period performance falls below a certain floor, while leaving control with the shareholders otherwise.

However, the solution is time inconsistent: there will be situations where the bank will be closed by the regulator even though its final period promises to be profitable. This model does illustrate one example in which forbearance is *ex post* inefficient but *ex ante* efficient. Banks are being closed "*pour encourager les autres*", and not because leaving them open is going to worsen their net position.

2.2 Inefficient liquidation: the **ex post** *problem*

Nevertheless, the dangers of allowing a bank with low or negative capital to operate are clear. The temptation is for shareholders to gamble on risky ventures with negative expected returns, given that they have little more to lose (the so-called deposit-put).

And excessive risk-taking by undercapitalized bankers may not be the worst that can happen, as is illustrated in Akerlof and Romer's (1993) model of looting. In this very

[4] The reputation and credibility of accounts even in advanced economies has taken several severe knocks in recent years, making Akerlof and Romer's (1993) complaints about inclusion of "goodwill" in the accounts of US Thrifts, and the transitory income-inflating effect of term transformation (when the yield curve is positively sloping) seem quaint these days.

simple model, bank management exploits limited liability, combined with the gullibility of depositors or the insouciance of the government as deposit insurer. It's not just a question of the management taking excessive risk in the hope that the value of the bank will improve. In Akerlof and Romer's model. Management seeks to extract as much value as possible (by legal or perhaps also illegal means) from the cash flow of a bank with the full intention of driving the bank into insolvency. As Akerlof and Romer note, when a bank moves into the looting phase, the "normal economics of maximizing economic value is replaced by the topsy-turvy economics of maximizing current extractable value, which tends to drive the firm's economic net worth deeply negative."[5] Only an undercapitalized bank will opt for a looting strategy. That is why this model is highly relevant for the analysis of forbearance. Indeed, it is the deliberate forbearance in the US allowing undercapitalized or insolvent intermediaries to function that motivated Akerlof and Romer's paper. Numerous developing country cases, from Mexico to the Philippines, though less thoroughly documented, seem illustrative of this behaviour.

Still, as exemplified in what Dewatripont and Tirole found in their three period model (mentioned above) the social costs that can be incurred by the process of closing a bank can militate against enforcement of capital adequacy requirements, and in effect represent a reason for forbearance (as in Sleet and Smith, 2000). For example, even though a bank's capital has fallen below the regulatory minimum, it may be holding illiquid assets which it would be socially inefficient to liquidate on closure. Likewise, the relationship capital that has been accumulated over the years by borrowers with their bankers could be lost.[6]

---

[5] In the real world, it is not always clear *ex post* whether loans going bad reflects high variance, or negative expected yields from the start. Akerlof and Romer suggest that a telltale indication of an bank that is being looted is the management's lack of concern about maintaining adequate documentation on loans: if the bank is "going for broke" the management will want those risky loans well-documented in order to be able to recover them if the gamble pays off. But if they are deliberately "going broke" it is the liquidator and not the current management that will have to attempt collection, as the current management do not expect to be around when the loans come up for collection. Looters will also seek nonbank collaborators to help them construct what will eventually be loss-making loan transactions but which spin off large cash flows in the short-run. The collaborators may be either naïve or unscrupulous; their involvement multiplies the social losses above the direct benefit to the looter. This deadweight cost of looting is exactly analogous to that demonstrated for corruption by Shleifer and Vishny (1993).

[6] On the other hand, cynics will point to the merits of breaking corrupt crony relationships between bank insiders and some of their preferred borrowers.

More generally, absent complete markets, it is possible for a tough regulatory enforcement to seem *ex ante* optimal, but to prove time inconsistent. The authorities don't want to close *this* bank, but only for fear of encouraging risky behaviour by other banks. [7]

With two- or three-period models, one can make a clear distinction between the *ex ante* and *ex post* problems. Things become a bit more blurred if we track the evolution of a weakening bank through time. As a bank nears its capital minimum, the risk of being closed soon affects the banker's incentive to conceal the true situation and can increase risk taking. In such circumstances, a stochastic closure rule can be better than a strict no-forbearance rule.

For example, Shim (2006) presents a multi-period model[8] in which the regulator cannot fully detect the profits being made by the banker nor the level of risk-reducing effort. The regulator sets deposit insurance premium and minimum capital levels to incentivize the banker to make the necessary efforts both to achieve profits and not to conceal them, even if the bank is close to closure. As a result, Shim finds that while prompt corrective action is appropriate, it should be applied stochastically: allowing bankers to hope they may be recapitalized, in order that may continue to make some efforts.[9] Here again some forbearance turns out to be optimal—but then again, how realistic is such a prescription for regulators in low-income environments.

This brief discussion of the literature on incentive effects—both *ex ante* and *ex post*—should be enough to show that these are sufficiently complex to have generated a prolonged and sophisticated literature which continues to generate conflicting

[7] In Mailath and Mester's (1994) paper, which also uses a three-period set-up, the regulator is not looking at capital adequacy, but instead at an indicator of whether the bank has adopted a risky or safe strategy in the first period. The decision as to whether to close comes in the second period. In this case too the inability of welfare-maximizing authorities to commit to a tough closure policy can result in higher risk-taking than would otherwise occur.

[8] The model is a rather abstract one, chosen in order to be able to exploit known results from stochastic control theory to derive the optimal program; however Shim shows that the model can be mapped to an implementation framework which has some comparability with real world banking regulation.

[9] In a slightly different set-up Kocherlakota and Shim (2007) also explore optimal regulatory policy, but in an environment dominated by collateralized lending. Instead of moral hazard, the volatility of collateral values is emphasized. If collateral values are not highly volatile, then they show that an optimal closure rule will involve some forbearance: conditions under which a bank that is surely going to impose some costs on the deposit insurer (because of the decline in posted collateral values), is nevertheless allowed to stay in business given the expected social returns of its future operations. If collateral values are very volatile, then forbearance will not be optimal.

theoretical conclusions regarding the optimality of forbearance, even when only one bank is being considered at a time. Additional complications arise because of the interdependence of banking firms.

2.3 Market interactions

In the real world there is more than one bank and the interaction of different banks also has a bearing on the consequences of regulatory closure decisions (cf. Goodhart et al., 2003, whose model also captures the linkage with monetary policy). One aspect here is the fragility represented by the possibility of a self-fulfilling depositor panic, analyzed for a single bank by Diamond and Dybvig (1983), and which underlies the fear that a disturbance in any part of the system could represent the trigger or coordinating event for a wider collapse. Such systemic collapses and their vulnerability to extrinsic events are modeled by Allen and Gale (2007, Chapter 5).

Linkages between banks on the lending side can crop up in several different relevant ways. For example, banks may make inferences about future prospects, including creditworthiness, that are based in part on looking over their shoulders at what other banks are doing (Honohan, 1999). Such information externalities are analyzed by a large and growing literature, (cf. Allen, Morris and Shin, 2006).

A contrasting form of interaction on the lending side arises with multi-bank lending, as well as with modern forms of structured asset-backed lending that also entail multiple lenders. These interactions have been modeled by Huang and Xu (2000) who integrate analysis of the interbank market with multi-bank project financing and the lender of last resort. Huang and Xu argue that having multiple financiers strengthens the financial system by imposing a harder budget constraint on borrowers. They assume that a consortium of financiers will find it difficult to agree on reorganization of a project in difficulty. [10] This will lead to socially inefficient closure of troubled projects, but the fact that a hard budget constraint is known to be in place will also improve the incentive for entrepreneurs to choose a better project in the first place. The lower average quality of projects, and the impossibility of detecting which

[10] The way in which the market for asset-backed securities dried-up in 2007 could be an example of this kind of situation.

projects are good and which are bad, results in a pooling equilibrium in the interbank market potentially resulting in a collapse of the market due to the lemons problem.[11]

Diamond and Rajan (2005) show that a form of contagion leading to systemic liquidity shortages may arise even without depositor panics or contractual links between banks. Their model requires banks to issue demandable debt (as a disciplining device on the bankers) which they use to finance potentially late-maturing projects. Banks can choose between liquidating or refinancing late-maturing projects, but this choice is not available to a bank which experiences so many late projects that it suffers a rational depositor run. Indeed, given the sequence of information flows, it may in this case have to liquidate even the early maturing projects. The consequence for systemic liquidity depends on the distribution of late projects between banks and thus on how many are run and have to liquidate their projects at a loss. It is a feature of their model that recapitalization (by the authorities) of illiquid failing banks may destroy may healthier banks and give rise to the need for massive recapitalization as the scramble for liquidity intensifies and projects are liquidated early. Instead, in their model, if the crisis cannot be resolved by a pure injection of liquidity into the system, recapitalization should be directed to the most liquid of the failing banks. More generally, they argue that intervention or forbearance policy needs to consider the general equilibrium, and not just the condition of each bank on its individual merits.

Conversion of a relatively isolated solvency shock into a systemic meltdown through collapse of the interbank market is, however, much less likely in an environment such as that in many developing countries where the interbank market is very small and inactive. This point is taken up in Section 4 below.

2.4 Regulatory incentives and behavior

Thus far, the discussion has implicitly assumed that regulators are choosing their actions in accordance with social welfare optimization. This assumption may not be valid. Different agencies are typically assigned more narrowly defined objectives

[11] Huang and Xu made the interesting suggestion that, when it comes to offering liquidity assistance to illiquid banks in the event of the pooling equilibrium leading to a collapse of the interbank market, the regulatory authorities should offer such assistance only at the price of a heavy rate of profits taxation. Offering such a contract can separate the solvent and illiquid from the insolvent, since the former will be reluctant to seek assistance. (Bagehot's penalty rate and equity instead of debt injections would be variants of this idea.)

(such as representing the interests of depositors, as suggested by Dewatripont and Tirole). Accordingly, another strand of the literature examines the incentives of the regulator, when several alternative institutions are present, such as the central bank as monetary authority, the deposit insurer as issuer of contingent quasi-fiscal liabilities and the prudential regulator may be evaluated and rewarded on the basis of success in protecting the interests of depositors. The questions of interest in this strand include the assignment of powers and responsibilities between different agencies, and the specification of the agency's explicit and implicit incentive structures. Indeed, as pointed out by Repullo (2000) and Kahn and Santos (2001), the deposit insurer and the central bank will experience a different impact on their profit and loss account in the case of closure. In Repullo's three-period model, depositor withdrawals in the middle period may leave the bank subject to closure. If it is to survive, the central bank has to make a loan, thereby putting its own money at risk. It stands to lose all of its loan if the bank fails (especially if, as is usual) the central bank's claim is junior to that of depositors. But the central bank's liability in the failure may be limited to the loan it has made. The deposit insurer, on the other hand, could be liable to meet all of the depositors' remaining claims if the bank fails in the final period. This illustrates the different incentives of the two institutions: the central bank may be more willing to keep the bank going than the deposit insurer. So it matters which of these two agencies has the responsibility to decide on closure.

Evidently, all of the analysis considered so far may be irrelevant if the regulator can be bribed for a small sum, or if the regulator's political instructions are to forbear on the capital adequacy problems of a troubled bank. These are problems more likely to be encountered in developing countries (Section 5). In either case stated regulations will impose little cost on a bank whose true capital has dipped below the regulatory minimum.

Theories of corruption (cf. Shleifer and Vishny, 1993) and incentive structures for bureaucracy (cf. Prendergast, 2006) have a clear potential application here. Perhaps the most relevant strand of theoretical literature here, though, is exemplified by Glaeser and Shleifer (2001), who argue the case for simple and easily verifiable regulations over the subtle and complex. This is essentially the same case as is made for bright lines over lengthy codes in other aspects of regulation. Bright lines mean lower verification and enforcement costs and greater ease of supervision of the regulatory agent by their principal. If the rules are clear, and verifiable, everyone can see when they are being violated: all market participants are called to witness and a new channel of market discipline is established. These benefits can be lost in the blurred environment of regulatory discretion and forbearance.

The literature on bright line standards in accounting presents similar conclusions in more elaborate models. For instance, Caplan and Kirschenheiter (2001) argue that use of "hard" bright line rules enhances the value of basic audits, limited to verification of facts, though it may do little to help things if a higher level of "expert" auditors are conducting the audits. It may well be that the sophistication and interpretive discretion of the typical bank supervisor could be classified as at "basic" rather than "expert".[12]

3. Information deficiencies in developing countries

Casual discussion of forbearance tends to assume that estimating the value of a bank's capital is a technical matter which can be done with a reasonable degree of confidence by well-trained regulators. The decision to forbear or not is, in this view, something which can be readily detected by all concerned. This is far from being the case in practice, however, especially in developing countries, and this fact must strongly influence the decision as to how much discretion to allow regulators in these countries.

The best way to illustrate the nature and extent of information problems is to illustrate by real world examples. We take five cases to illustrate the kinds of problem that arise.

First, there is the case of the Egyptian state-owned banks. After 2002 it was recognized that the four large state-owned banks that dominate Egypt's banking system needed to be restructured and recapitalized, a process which is still under way,

[12] As the accounting literature observes, very precise rules can create precise "safe havens" for malpractice; and this could present problems for bank accounting too (Nelson, 2003).

with one of them having been privatized in 2006. Undercapitalization was not indicated in the audited accounts of these banks; indeed, average capitalization of the banking system was about 10 per cent. The interesting thing, though, is the evidence that, from 1994 to 2001, bank management were likely aware at least partially, that the accounts they were presenting were optimistic. Appendix 1 shows the evidence for a pattern of profit reporting that strongly suggests a strategy of limiting loan loss provisioning to an amount that still left reported profits slightly positive. This example of upside-down accounting carried out over an extended period by some of the largest banks on the African continent is a clear indication that apparent compliance with capitalization requirements is no guarantee of actual compliance, even when some agents in the system know that the accounts are misleading.

A second case is the Chinese banks. There are similarities here with the case of Egypt in that it concerns State-owned banks. Action by the authorities to recapitalize these banks has been ongoing since 1998. A complicated series of transactions had resulted by 2006 in net fiscal costs in excess of USD 300 billion equivalent—the largest banking bail-out in history (Honohan, 2007; Podpiera, 2005). This amount, which is unlikely to be the full account even for the four main state-owned commercial banks, does not include fiscal costs relating to other parts of the banking and nearbanking system, which could bring the grand total close to USD 500 billion – or about 50 per cent of reported 2001 GDP. Most of these transactions have been in the form of asset purchases by other state agencies at prices that have proved to be well above recoverable values: the subsidy embedded in these transactions has not been officially acknowledged. To what extent the ultimate cost was known, and to what extent the need for an injection of funds on this scale only dawned gradually on regulators is hard to assess. Lardy (1998) set out a plausible account of how the banks had been deliberately used as an alternative to the use of artificial prices and direct transfers to support some state-owned enterprises in the process of China's transition to a market economy. He foresaw significant deferred fiscal costs in this practice. A wide range of analysts' forecasts for total losses was presented during the years 1998 to 2006; these generally presented a wide range or merely confined themselves to general statements such as that in the Fitch report of May 2002, to the effect that "in practice, the banking system is substantially insolvent".

With such a wide range of estimates (even if some of the relevant regulatory authorities may have had more precise information – though with incoherent governance structures at the banks in this period that assumption cannot be made) it is easy both to argue that this has been a large and protracted case of capital adequacy forbearance and at the same time to wonder whether the authorities really could have defined within any acceptable margin of error what recapitalization would be required to bring the banks up to regulatory standards.

Instead of attempting a fully realistic calculation of the banks' capital and insisting that it should be brought up to regulatory minima, the authorities embarked on a multi-stranded decade-long effort to put the banks' finances to rights. The negative capital position was never explicitly acknowledged, but it was corrected through a variety of restructuring mechanisms, mainly loan sales to newly created asset management companies (shifting the measurement issues to these nonbanks, while improving the precision of the estimate of capital of the banks). At the same time, the authorities also placed pressure on bank managements not through capital targets but instead by setting and enforcing (from 1998) a target for the proportion of the loan portfolio that was non-performing. Each of the four state-owned banks was required to lower this proportion by 2-3 percentage points per year. Top management of the banks were made accountable for achieving this target which was given a high profile. Indeed, judging from the annual reports of the banks, it appears to have been met. But NPL ratios, while readily measured, can be improved without achieving a substantive improvement in solvency prospects. Writing-off an NPL, or rescheduling the loan, are just two ways of lowering this ratio without improving the capitalization of the bank. That, despite these shortcomings, the Chinese authorities should have focused on NPLs rather than capital points to their recognition of the severe measurement difficulties with capital adequacy itself.

(It may reasonably be asked whether a state-owned bank needs to be held to a high level of capital. Without private shareholders, the adverse incentive effects of low capitalization are not so clear, and the role of capital as a buffer is also moot: the capital is owned by the state, so its loss entails the same fiscal cost as would be involved in compensating losses incurred by a zero-capital bank. Curiously, this point is not widely accepted in the policy literature. In the Chinese case, it has been superseded by the decision to part-privatize the banks).

The case of Lebanon is more complex still. The Lebanese banks have sizable claims on the Government of Lebanon, which indeed is more heavily indebted to the banking system (relative to GDP) than any other government in the world. Lebanese banks also attract an extraordinarily high volume of deposits in relation to GDP, and over two-thirds of these are denominated in US Dollars reflecting the high dollarization of the Lebanese economy generally as well as the sizable openness of that economy. The fiscal situation in the Lebanon has seemed unsustainable to many observers at different points over the past decade or more, as has the currency peg, which has been held for a decade. The banks' claims on the government include some in local currency and some denominated in USD. The Lebanese banks cannot survive a collapse of the currency and/or the market value of government debt. How should these sizable risks be factored into the accounting for Lebanon's banks? At present, no special provisions are being made, and the accounts can be seen as fair and accurate conditional on survival of the currency and fiscal situation, but perhaps not otherwise. In practice, however, on two occasions in the past 5 years, just when default seemed increasingly unavoidable, a special donor conference has coughed up sizable transfers and loans to Lebanon allowing the evil day to be postponed. It is not inconceivable that this pattern could be repeated indefinitely. So what is the true value of the Lebanese banks' portfolio and are they to be considered truly adequately capitalized? The difficulty of answering this question highlights the questionable precision of bank accounts in developing countries generally, even if the measurement problems are not so acute or indeed endogenous to the health of the banking system.

We will pass over the numerous cases of other banking systems which were overwhelmed by macroeconomic downturns and where *ex post* the degree to which the final deteriorated condition of the banks owes more to the economic downturn and how much to flawed underwriting (or 'crony capitalism'). These too point to the difficulty of providing a reasonable range for the ex ante underlying value of each bank's capital.

There have also been several large individual bank failures in developing countries, often associated with fraud, which were so large as to have systemic ramifications. Perhaps the cleanest example of this, and a case where forbearance over a period of several months during which large liquidity loans from the central bank were looted by bank insiders, was that of Banco Latino in Venezuela 1994-95. The regulator was not equipped with adequate legislative support at the time of this crisis (thanks to parliamentary opposition likely fomented or at least abetted by looting bank owners). However, it appears that the regulator did not consider Banco Latino to be insolvent, yet allowed liquidity loans amounting to USD 9 billion (or about 15 per cent of Venezuela's GDP) to be lost due to inadequate information about the bank's balance sheet.[13]

Similar in magnitude, the failure of Banco Intercontinental (Baninter), the third largest bank in the Dominican Republic in May 2003 also revealed a huge accounting fraud. Here was a case of the diverted deposits fraud (Caprio and Honohan, 2005) where the regulator is shown a set of accounts which omits a segment[14] of the deposits which have been placed with the bank, and which are being looted by insiders. Given the scale of the discrepancy in the case of Baninter (reportedly USD 2.2 billion equivalent to about a third of the end-2002 deposits in the banking system, or about 11 per cent of GDP; liquidity loans to Baninter and two other weak banks hit by contagious depositor withdrawals totaled more than twice that sum) it is hard to understand how the regulator could not have suspected that something was amiss.[15,16]

Some have held out the hope that private rating agencies would have the necessary incentive and market information to do as well or better than the official regulator in sniffing out problematic sets of accounts, with a focus on default risks. An increasing number of banks in developing countries are being rated by the international rating

[13] More than one bank was involved in this rather complex story. In essence, it appears to have been a version of the diverted deposits fraud discussed later.
[14] As much as 90 per cent in the case of Imar Bank in Turkey in 2003 (Soral et al., 2006)
[15] An IMF-World Bank FSAP assessment was carried out shortly before the collapse but did not pinpoint the failing bank or foresee the crisis.
[16] Brownbridge (2002) describes the failure of Greenland Bank, Uganda. At June 1998, the managers were reporting the bank as solvent, but a special audit conducted by a big-5 international accounting firm found a net capital deficiency of USh 0.3 billion, or about 0.3% of the bank's total assets. However, the audit and investigation conducted after its closure just nine months later in April 1999 found unreported deposits and assets and estimated a negative net worth of over UGX 62 billion, or about 60 per cent of total assets.

firms (Caprio and Honohan, 2004). However, they too have failed to anticipate large bank collapses, even those that were unrelated to macroeconomic downturns. For example, the failure in mid 1995 of Banco Nacional and Banco Economico in Brazil, with deficiencies totaling almost USD 10 billion, were not foreseen by international analysts.[17] And a December 2002 Fitch report on the Dominican Banking System noted robust profitability and struck no note of caution, this just five months before the devastating banking collapse in that country already mentioned above.

The examples provided show that precise information is often not available to regulators—or other market participants—in developing countries. But this is not because of structural complexity in these markets; indeed, developing country banking systems display much less interconnectedness than do those of advanced economies.

4: Interconnectedness, contagion and liquidity

Interconnectedness is less in developing country financial systems. This arguable proposition, with its implication that a disturbance to one intermediary is less likely to be transmitted to others, is based on three distinct observations. First, interbank markets are typically small in relation to the size of the banking system and relatively inactive. Second, the financial systems are small relative to the economy. Third, large depositors and many of the larger firms have access to international financial markets. These considerations do not rule out the emergence of a contagious event spreading illiquidity through the system at large, but they reduce its likelihood.

Interbank markets tend to be smaller and less active partly because of the limited diversity of banking activities, partly because of the limited use of derivatives and the

[17] Neither of these Brazilian failures could easily have been foreseen from the published data. In particular, excessive asset growth was hardly evident: thus, although Nacional's market share (among the five largest private banks) had increased from 14 to 26 per cent between 1990 and 1993, it had since fallen back; Economico's market share had slipped in the five years before its failure. Furthermore, both banks were reporting a slightly lower share of non-performing loans than their peers. Admittedly, Nacional's reported risk-weighted capital adequacy was, at about 9 per cent, well below that of the other large private banks, its shares were trading on a low price-earnings ratio, and other banks were no longer lending to it on the interbank market. But as late as November 1994 international analysts Salomon Brothers described it as a "strong" bank which they believed would be a "long-term winner"; and in June 1995, less than five months before the deposit run that precipitated the bank's failure, they recommended the shares as a "hold".

smaller volume of payments transactions related to securities trading, and partly because of the lower credit ratings of many banks in these countries. Indeed, the pattern in many countries is for the bulk of the interbank borrowing to be made by the local branches or subsidiaries of international banks who may have a smaller retail franchise and less direct access to local currency deposits.[18]

When interbank markets are small and the net takers are (in the event that they are unable to rollover their interbank borrowings) in a position to finance themselves from parent companies abroad, the likelihood of many types of events studied in the literature on liquidity crises is greatly reduced.

The small size of many financial systems also reflects the degree to which the economy is not inextricably permeated with contracts that depend on continuous smooth functioning of the financial system. Of course the widespread collapses of retail markets in Argentina in the crisis of 2001-2 does show how basic economic functioning does require the banking system on a continuous basis. On the other hand, the survival of the BRI village microfinance system through the 1998 banking crisis of that country displays the degree to which even some financial intermediation can survive the widespread failure of banks. The rapid recovery (helped by energy prices) of the Russian economy also points in the same direction. And the credit-less "phoenix" recoveries documented by Calvo et al., (2007) also show that the contribution of financial depth to growth is not simply to be measured by short-term fluctuations in credit in such economies.

Finally, all developing country systems are small relative to global finance and in many countries there are appreciable links between the larger economic agents and the rest of the world. On the deposit side these are graphically illustrated by Figure 1 which shows the ratio of offshore bank deposits held by residents to domestic bank deposits. Although the phenomenon is strongest in Africa and Latin America, this reserve pool of liquidity (as well as international credit lines that are available to the

[18] This is especially noticeable in the many countries where foreign currency deposits have begun to take a sizable fraction of the deposit market. Often it is the foreign banks that get a disproportionate share of the foreign currency deposits but are short of local currency to lend to their corporate clients.

larger firms) is appreciable and also makes the likelihood of the kinds of liquidity crunch described in several models smaller.

This is not to say that liquidity crises are impossible. Far from it. But many of the liquidity crises that have been seen are closely related to fears of exchange rate collapse or of default by national authorities—not to self-generated failures arising in the banking system.

The main type of bank-related liquidity crisis that these remarks do not speak to is depositor panic associated with a contagious reassessment of the solvency and liquidity of banks in general following revelation (whether from regulatory action or otherwise) of solvency problems somewhere in the system. There have undoubtedly been a number of such events in recent years.

The depositor response to the closure of banks in Indonesia is one such case.[19] The closure of 16 banks, accounting for about 3 per cent of the system, in Indonesia in October 1997 is often cited as an example of how lack of forbearance can trigger a panic. In fact a closer look at the situation supports quite the opposite view. It is acknowledged that the closure event heightened depositor uncertainty and contributed to considerable withdrawals over the following months. But these depositor withdrawals were not irrational responses that caused other banks to fail. Instead, they were at least in part a response to the partial nature of the policy intervention, which left open most of the weak banks—almost all of them much larger than the closed 16—without indicating whether and on what basis there would be further closures. No wonder that depositors felt sure that the closures were only the start (audits made in mid 1998 confirmed that insolvency in the banking system was already entrenched by late 1997, cf. Enoch et al., 2001). The scale of deposit withdrawals fed, and was exacerbated by, a collapse of the exchange rate, worsening the bank insolvency. The authorities decided in January 1998 to guarantee the full amount of bank deposits. By March of the following year a total of 79 of the 222

[19] Kenyan depositors also ran from a class of banks that began to have difficulties in the late 1990s—though in this case too it seems that there were widespread problems in the class of banks affected by the runs.

banks in Indonesia (accounting for over 40 per cent of the total assets of the system) had been closed, merged or nationalized. (Fane and McLeod, 2002).

So while the handling of the banking weaknesses by means of the initial closure was highly destabilizing, and while the subsequent collapse of the exchange rate undoubtedly aggravated the insolvency of the remaining banks (Radelet and Sachs, 1998), it seems likely that a more comprehensive intervention into weak banks (thus *less* forbearance), accompanied by a clear statement of future closure policy, and combined with a coherent macroeconomic strategy would have been better.

It might seem perverse to question the importance or frequency of contagion within developing country banking systems when so many systemic events have occurred. The point is, though, that what appears to be contagion is so often a common cause (e.g. a macro boom and bust cycle) or a simultaneous uncovering of a widespread deficiency (Honohan, 2000). Quite often, the denouement occurs after a change of government exposes a pattern of politically motivated forbearance, and this points to the pervasive agency problems that are so prominent in banking regulation in much of the developing world.

Section 5: The double agency problem and the goals of regulators

The considerable recent investment in most developing countries in upgrading regulatory capacity cannot be assumed to have overcome the considerable agency problems that still exist. It is true that there has been a fairly direct institutional and policy transplantation of advanced country practice in the developing world over the past two decades. A process of regulatory convergence has occurred, at least on paper, assisted in the past decade by the Basel Core Principles and the formalized assessments of national compliance with these Principles carried out mostly under the auspices of the International Monetary Fund and the World Bank.

Yet even if the forms and even the legal powers are converging, account has to be taken of the differing nature and intensity of incentive and agency problems that surround the implementation of these regulatory structures. Even in advanced economies enjoying a sophisticated electorate and a free press, as well as relatively efficient and impartial law enforcement, it is evident that regulators are subject to pressures and incentives that make it unrealistic to suppose that they are single-mindedly pursuing the public good.

Combining data on reported corruption in banking across countries, with the Barth et al. (2006) data on the structure and style of banking regulation, Beck et al. (2005) made the remarkable discovery that banking corruption was significantly correlated with the degree of regulatory discretion. The conclusion drawn by these authors is not only that granting bank regulators discretion creates rents which will be manifested in corrupt transactions between regulator and bank, but also that somehow a culture of corruption will effectively permeate the entire banking system leading to the need for borrowers to bribe bank officials and so on.

Of course it is realistic to think of the regulator as an agent of a higher level of government. To an extent, those who work as regulators may be self-selected by a desire to ensure safe and sound banking, or can become indoctrinated in a favorable regulatory culture. However, this does not mean that that the regulator will always pursue the public interest effectively. For one thing, uncovering errors, misjudgments and fraud in banking is not a simple task and requires skill, experience and assiduity. These qualities may not be sufficiently present in the regulatory authority.

But in addition, there are reasons why, even if problems are detected, enforcement is weak. This can be because delinquent insiders at regulated banks are prepared to bribe the regulator or because the regulator fears the consequences of enforcement action either at a personal or an institutional level. Despite efforts in most countries to protect individual regulators from being sued for carrying out their work in good faith, it seems to be astonishingly difficult to provide watertight protection in this regard. A decision to enforce regulatory action can mean years under the shadow of court proceedings for an individual regulator. In only 20 of 149 countries surveyed by Barth et al. are banks unable to appeal to courts against a decision of the supervisor.[20]

[20] Even in the UK, where regulators enjoy a high degree of protection, regulatory action or inaction can result in lengthy litigation. The unsuccessful case taken by the liquidators of the BCCI against the Bank of England concluded only in 2005, almost 15 years after the failure of BCCI. Retired regulators in their 70s had to appear in court to defend their actions – or in this case inaction.

Besides, the delinquent bankers may themselves have political protection, being part of an integrated elite which embraces political, financial and industrial dimensions (as discussed in the wider context of financial globalization by Stulz, 2005).

Regulators who covertly forbear may never be found out. Although in some instances forbearance is explicitly acknowledged and documented,[21] it is often extraordinarily difficult to detect deliberate under-enforcement.[22] Under-enforcement will often not show up immediately in bank failure, as banks can often survive for years in insolvency. There rarely is an audit trail establishing who knew what when. And when the bank failure is eventually detected, the original under-enforcer is unlikely to be still in charge of the desk. For one thing, the failure often emerges only after a change of political regime, when the relevant bankers have lost their political protection (and the chief regulator may have lost their job also). Besides, the failure will often become evident only during a more general economic downturn, and may often be blamed on that downturn, rather than on management failings of several years before.

Several agencies are typically involved in bank closure decisions. Although the banking supervisor may seem to be in the frontline of the decision, the big decisions are often taken elsewhere. Among the reasons for this are the fact that the banking regulator usually does not have the financial resources to meet liquidity needs of an illiquid bank[23] and that (despite the recommendation of the Basel Core Principles) the banking regulator's closure decisions are often subject to a political override.[24]

In fact, the key role of last resort lending by the central bank in several recent failures—most spectacularly in the Venezuelan and Dominican cases mentioned

[21] As in the Turkish case discussed by Soral et al. (2006)—though even then, the bank concerned appears to have been concealing much fraud from the supervisor.
[22] The length of the unsuccessful attempts to prove "misfeasance" (a variant of negligence) by UK regulators in the case of the BCCI illustrate the problem even in an advanced economy context. In contrast, Shleifer and Vishny (1993) suggest that a good accounting system can often detect other forms of corruption such as, for example, the granting of licenses for a bribe to the responsible official instead of the license fee.
[23] The insolvency of the FSLIC has been mentioned as one of the reasons for excessive forbearance in the US S&L Crisis (cf. Black cited in Akerlof and Romer, 1993, p.68)
[24] India, Italy, New Zealand and francophone countries in Africa are among those where the law explicitly states that decision on closure is taken by a Government Ministry and not by an independent specific financial regulator (cf. World Bank Regulatory Survey).

above—highlights the fact that it is this agency which by default makes the decisions re closure by extending (or less often by not extending) liquidity loans to a troubled bank while the regulatory agency and deposit insurer stand by.

6 Interaction of information and agency problems

Do the heightened information and incentive problems in developing countries argue for or against forbearance? A simple model allows some of the issues to be disentangled (Appendix 2). It suggests the following intuitive argument.

Heightened uncertainty and information deficiencies increase the risk of bank failure for any given level of enforced capital adequacy. Given that the regulator may only have an uncertain estimate of the bank's capital, enforcing a certain level of measured capital means an increased probability that a bank is being allowed to operate with less capital than envisaged, which increases the risk of failure because the cushion of capital against unexpected loan and other losses is higher, and probably also because risk-taking at lower capital levels may be higher by the usual limited liability argument.[25]

Thus, even if the regulator does not know how much capital the bank has, is this not an *a priori* reason for being flexible in terms of enforcing any given capital adequacy requirement? Being flexible means tolerating lower capital adequacy and in effect applying a lower capital adequacy standard. But if the risk of failure is higher for any given level of enforced adequacy, this is adding even more to the risks.

Given the pressure (from the regulated institutions) on regulators to forbear, i.e. not to require a capital injection into a bank thought to be undercapitalized, it is likely that increased uncertainty will in practice result in even more forbearance. The regulator's decision on forbearance will be based on balancing the pressures to forbear against the risks of a failure which may be blamed on the regulator. A relaxation of standards will increase the risk of failure but will also respond to the pressure from bank insiders. If what matters to the regulator is just the risk of bank failure, then the

[25] Indeed, as uncertainty increases, the risk of failure increases more rapidly with falling levels of enforced capital adequacy.

regulator *may* be induced by higher uncertainly to increase the enforced level of capital adequacy. However, this stiffening effect is reduced and may be reversed if the public is unable to detect forbearance.

It is in this final point that the argument for a regime which inhibits forbearance lies. The ability of the regulator to conceal forbearance that is motivated by political pressures or the private benefit of the regulator can be inhibited by a bright line regulatory regime. According to the line of argument developed, this will result in decisions that are closer to the public interest.

If there are good policy reasons for forbearance in a particular instance, these can still be put forward in justification of such action

7 Seeking empirical evidence

This Section reviews existing cross-country empirical evidence on the impact of forbearance on banking crisis costs, and discusses the difficulties that exist in extending it to answer the question posed by the present paper. There is little convincing empirical evidence, one way or the other, on the effectiveness or risks of a policy of forbearance. It is hard even to obtain an empirical counterpart to the concept of forbearance itself. And as mentioned there can be degrees of forbearance.

The US case is a leading one. The prolonged period of forbearance during the US Savings and Loan crisis of the 1980s was held by many scholars to have contributed greatly to the scale and cost of this crisis. A consequence of the ensuing debate in the US was the enactment in 1991 of FDICIA, which inter alia mandates on regulators prompt corrective action where the banks and nearbanks under their remit become undercapitalized.[26] Even if they do not amount to 'zero tolerance', the limitations on regulatory discretion under FDICIA appear to have reduced the amount of forbearance. The sharp drop in the number of institutions which have had to be intervened since FDICIA were taken by enthusiasts as an indicator of the effectiveness of prompt corrective action under the new Act. Thus, the percentage of

[26] With this measured at market value wherever possible. The market valuation approach introduces a degree of procyclicality as indicated below if asset liquidations in the crisis depress market prices.

undercapitalized banks in the US fell from 5 per cent in 1990 and 3 per cent in 1991 to about 0.2 per cent by 1996 (Benston and Kaufman, 1997). And there is evidence that supervisors have acted more quickly and more effectively to bring faltering banks back to health since the enactment of FDICIA (Kane et al. 2007). Confirming that the restrictions on forbearance in the Act are causal in this regard is less easy. (And not all of the evidence favours the interpretation that FDICIA has tightened forbearance as much as intended. For example, Hanweck and Spellman point to evidence from the market for subordinated debt, that market participants anticipate longer periods of forbearance than the maximum of 270 days envisaged in FDICIA.) But above all, it would be unwise to extrapolate experience from the US to the case of the developing countries for the reasons explored in Sections 3-5.

Cross-country experience with financial crises can be used to throw some empirical light relevant to developing countries. Looking at the fiscal costs associated with some 40 crises for which data were available, Honohan and Klingebiel (2003) found that capital forbearance (and other accommodating policies) tended to be statistically associated with higher fiscal costs (as a percentage of GDP). Although the sample covered a very wide range of income levels, their study did not attempt to assess whether these effects were more or less evident for low income countries. Table 1 show the results from extending the analysis in this way. Per capita GDP (measured at purchasing power parities) is interacted with each of the regulatory policies and added to a representative core regression in HK (2003). The interaction terms are not together statistically significant (regression B), and an OLS regression strategy approach ends with either regression C or D, which suggests that the impact of capital forbearance on fiscal costs is not income-sensitive, though use of loan guarantee schemes may be a relatively less costly crisis-response policy in higher income countries. Because of the possible endogeneity, two-stage least squares estimates are also reported. These suggest that even the interaction term with loan guarantees is actually insignificant. This dataset thus provides little basis for asserting that income levels matter for optimal policy.

It is worth stressing a methodological shortcoming of the Honohan-Klingebiel approach, already hinted at in their paper, namely the selection bias entailed in running regressions that only include crisis events. What of the instances when forbearance may have prevented a crisis? Such episodes are not included in the sample. In effect, what is being estimated is the loss conditional on there being a crisis. Whether adopting an accommodating policy might also reduce the probability of having a crisis is not so clearly addressed by this data set or this study. To be sure, the authors did include some macroeconomic determinants of the crisis (θ=1), but still there is a sample selection issue given that the variable of interest is $E[y]$, where

$$E[y] = \Pr[\theta = 1]E[y|\theta = 1].$$

Lacking a strong selection equation for the determinants of θ=1 (Honohan, 2000), and equally importantly, lacking data on policies adopted in non-crisis situations, we have to remain somewhat agnostic about the overall impact of policy on $E[y]$.

To an extent, the degree of forbearance is something that is detected only in practice. Untested regulatory policy is likely to be more rigid in theory than its implementation will be in practice. Nevertheless stated regulatory policy may be somewhat informative.

Seeking to explain differences across countries in bank ratings, seen as an indicator of systemic banking risk, Demirgűç-Kunt, Detragiache and Tressel (2006) used unpublished assessments of national compliance with various elements of the Basel Core Principles for Banking Supervision. They found that the only parts of the Core Principles that were significantly correlated with bank ratings were the chapter referring to availability and publication of financial information,[27] and that referring to licensing powers. In particular, that part of the Core Principles most likely to indicate forbearance, namely a country's score on the chapter relating to enforcement powers of regulators does not help to predict its banks' average ratings.

Power of regulators can refer not only to their capacity to enforce regulations, but to their discretion in doing so. Barth et al. (2006) assembled data on regulatory style and procedures and argued that increasing supervisory "power" was often counterproductive, in contrast to regulatory measures that increase the information

[27] Perhaps it is not surprising that information transparency is particularly favored by rating agencies.

and incentives for private sector monitoring. If power and forbearance were opposites, then this might suggest that forbearance was a good thing. But the composition of the "power" variable is not simple and needs to be carefully considered in this context. It is formed from the answers to no fewer than 14 questions, ranging from whether auditors are required to report off-balance sheet exposures to the supervisor to whether the supervisor can suspend the director's decision to distribute management fees. It does *not* include either of the questions on whether supervisors have discretion to forebear. Interestingly, the degree of supervisory discretion[28] is negatively, albeit not strongly, correlated with supervisory "power" (R= –0.21).

In order to assess the impact of regulation on the incidence of banking crises, Barth et al. regressed an indicator function of a systemic crisis 1988-98 on their indices of regulatory practice (as of 1999-2001). The "power" variable is not significant. The forbearance variables are not included in the reported regressions.[29]

Most of the empirical evidence does not, therefore speak very clearly on the issue at hand. What evidence there is seems to argue against the exercise of supervisory discretion, but it is hardly decisive.

8. Concluding remarks

The degree of discretion to be allowed in regulatory design, and when that discretion should be exercised, will remain one of the most contentious issues in bank regulation. Theoretical models cannot decide the issue. And the prospect of a decisive econometric contribution that would decide the matter seems remote. But even if advanced countries can surely benefit from the judicious application of some forbearance, developing countries face severe difficulties of agency and of information which argue for less regulatory discretion.

[28] They also have an index of prompt corrective action which has a reassuringly strong negative correlation (-0.73) with the forbearance index.

[29] As they acknowledge, it is not clear how to interpret a regression purporting to explain crisis incidence in the 1990s by regulatory practice at the end of that decade. Nevertheless, as a robustness check to the earlier work, we may add the Barth et al. indices of regulatory practice to the Honohan-Klingebiel set of explanatory variables. As can be seen from Table 2, the selected Barth et al. indices are not significant, and they do not change the sign of the HK explanatory variables.

If regulators cannot be relied upon to pursue to public interest effectively, it may be better to limit their forbearance discretion, allowing the bright line of the basic rules to be more easily and visibly enforced.

The dramatically worse information and accounting environment does mean that any mechanical rule is unlikely to be fed with good information, but it also reduces the regulators' exposure to criticism inasmuch as forbearance can more easily be concealed. Thus the information deficiencies also argue for less forbearance discretion.

Finally, the more limited interdependence of financial sector intermediaries in developing countries suggests that the risks of unwittingly triggering a chain reaction through a fragile system are less.

To be sure, an absolutist zero tolerance approach would hardly be defensible. Instead, measures to strengthen accountability and transparency of regulatory action (perhaps with an appropriate time-lag) should be adopted, together with a tightening of capital requirements to take account of the accounting uncertainties.

Enforcement need not be mechanical, but there should be a stronger presumption of enforcement in developing countries, especially where agency and governance problems are thought to be severe.

Bibliography

Akerlof, George A. and Paul M. Romer. 1993. "Looting: The Economic Underworld of Bankruptcy for Profit." *Brookings Papers on Economic Activity* 2:1-73.

Allen, Linda and Anthony Saunders. 1993. "Forbearance and Valuation of Deposit Insurance as a Callable Put." *Journal of Banking and Finance* 17(4): 629-643.

Allen, Franklin and Douglas Gale. 2007. *Understanding Financial Crises*. Oxford University Press.

Allen, Franklin, Stephen Morris and Hyun Song Shin. 2006. "Beauty Contests and Iterated Expectations in Asset Markets." *Review of Financial Studies* 19:719-752.

Barth, James R., Gerard Caprio, Jr. and Ross Levine. 2006. *Rethinking Bank Regulation: Tills Angel Govern*. New York: Cambridge University Press.

Beck, Thorsten, Aslı Demirgűç-Kunt, Ross E. Levine. 2005."Bank Supervision and Corruption in Lending." NBER Working Paper 11498.

Benston; George J. and George G. Kaufman. 1997. "FDICIA After Five Years." *The Journal of Economic Perspectives*, Vol. 11, No. 3. (Summer, 1997), pp. 139-158.

Brownbridge, Martin. 2002. "Resolving Bank Failures in Uganda: Policy Lessons from Recent Bank Failures." *Development Policy Review* 20(3):279-91.

Calvo, Guillermo A., Alejandro Izquierdo and Ernesto Talvi. 2006. "Phoenix Miracles in Emerging Markets: Recovering without Credit from Systemic Financial Crises." NBER Working Paper No. 12101.

Caplan, Dennis and Michael Kirschenheiter. 2004. "A Model of Auditing under Bright-Line Accounting Standards." *Journal of Accounting Auditing and Finance* 19(4): 523-564.

Caprio, Gerard and Patrick Honohan. 2004. "Can the Unsophisticated Market Provide Discipline?" in William C. Hunter, George G. Kaufman, Claudio Borio and Kostas Tsatsaronis, eds. *Market Discipline across Countries and Industries*. Cambridge, MA: MIT Press. Pp. 349-362.

Caprio, Gerard and Patrick Honohan. 2005. "Starting Over Safely: Rebuilding Banking Systems." in Gerard Caprio, James A. Hanson and Robert E. Litan, eds., *Financial Crises: Lessons from the Past, Preparation for the Future*, (Washington DC: Brookings Institution Press), pp. 217-255.

Demirgüç-Kunt, Aslı, Enrica Detragiache, and Thierry Tressel. 2006. "Banking on the Principles: Compliance with Basel Core Principles and Bank Soundness." *IMF Working Paper* 06/242.

De Juan, Aristobulo. 2002. "From Good Bankers to Bad Bankers." In Gerard Caprio, Patrick Honohan and Dimitri Vittas, eds. *Financial Sector Policy for Developing Countries*. New York: Oxford University Press.

Dewatripont, Mathias and Jean Tirole. 1994. *The Prudential Regulation of Banks*. (Cambridge, MA: The MIT Press).

Diamond, Douglas W. and Philip H. Dybvig. 1983. "Bank Runs, Deposit Insurance and Liquidity." *Journal of Political Economy* 91:401-419.

Diamond, Douglas and Raghuram G. Rajan. 2005. "Liquidity Shortages and Banking Crises." *Journal of Finance* 60(2): 615-647.

Elizalde, Abel and Rafael Repullo. 2006. "Economic and Regulatory Capital in Banking: What is the Difference?" Working paper.
ftp://ftp.cemfi.es/pdf/papers/repullo/economic%20capital.pdf

Enoch, Charles, Baldwin, Barbara; Frécaut, Olivier and Kovanen Arto. 2001 "Indonesia: Anatomy of a Banking Crisis Two Years of Living Dangerously 1997–99."
IMF *Working Paper* WP/01/52.

Fane, George and Ross H. McLeod. 2002. "Banking Collapse and Restructuring in Indonesia, 1997-2001." *Cato Journal*, 22(2):277-195.

Glaeser Edward L. and Andrei Shleifer. 2001. "A Case for Quantity Regulation." The *American Economic Review*, 91(2): 431-435.

Goodhart, C.A.E. 2007. "Is Forbearance Always Bad? How Can We Test Whether Its Use Lessened the Incidence of Crises?" *Central Banking* forthcoming.

Goodhart, C.A.E., P. Sunirand and D.P. Tsomocos. 2006. "A Model to Analyze Financial Fragility." *Economic Theory* 27: 107-142.

Hanweck, Gerald A. and Lewis J. Spellman. 2002. "Forbearance Expectations and the Subordinated Signal of Bank Insolvency." Mimeo.

Honohan, Patrick. 1999. "A Model of Bank Contagion Through Lending." *International Review of Economics and Finance*, 8(2): 147-163.

Honohan, Patrick. 2000. "Banking System Failures in Developing and Transition Countries: Diagnosis and Prediction." *Economic Notes* 29(1): 83-109.

Honohan, Patrick and Daniela Klingebiel. 2003. "The Fiscal Cost Implications of an Accommodating Approach to Banking Crises." *Journal of Banking and Finance* 27(8): 1539-1560.

Honohan, Patrick. 2007. "Protecting Depositors in China: Experience and Evolving Policy", in Aslı Demirgűç-Kunt, Edward Kane and Luc Laeven, eds., *Deposit Insurance around the World: Issues of Design and Implementation* Cambridge, MA: MIT Press. 2007 Forthcoming.

Huang, Haizhou and Changgang Xu. 2000. "Financial Institutions, Financial Contagion and Financial Crises." *IMF Working Paper* 00/92.

Kane, Edward J., Rosalind L. Bennett and Robert C. Oshinsky. 2007. "Evidence of Improved Monitoring and Insolvency Resolution after FDICIA." Mimeo.

Keeley, M.C. and F.T. Furlong. 1990. "A Re-examination of Mean-Variance Analysis of Bank Capital Regulation." *Journal of Banking and Finance* 14: 69-84.

Kerkhof, J. and B. Melenberg. 2004. "Backtesting for Risk-based Regulatory Capital." *Journal of Banking and Finance* 28(8): 1845-1866.

Kahn, Charles M. and Joao A.C. Santos. 2001. "Allocating Bank Regulatory Powers: Lender of Last Resort, Deposit Insurance and Sueprvision." BIS Working Paper 102..

Kocherlakota, Narayana and Ilhyock Shim. 2007. "Forbearance and Prompt Corrective Action." *Journal of Money, Credit and Banking* 39(5): 1107-29.

Kopecky Kenneth J. and David VanHoose. 2006. "Capital Regulation, Heterogeneous Monitoring Costs and Aggregate Loan Quality." *Journal of Banking and Finance* 30: 2235-2255.

Lardy, Nicholas R. 1998. *China's Unfinished Economic Revolution*. Washington, DC: Brookings Institution.

Nelson, Mark W. 2003. "Behavioral Evidence on the Effects of Principles- and Rules-Based Standards." *Accounting Horizons* 17(1): 91-104.

Peura, Samu and Esa Jokivuolle. 2004. "Simulation Based Stress Tests of Banks' Regulatory Capital Adequacy." *Journal of Banking and Finance* 28:1801-1824.

Podpiera, Richard. 2006. Progress in China's Banking Sector Reform: Has Bank Behavior Changed?" *IMF Working Paper* 06/71.

Radelet, Steven and Jeffrey Sachs. 1998. "The East Asian Financial Crisis: Diagnosis, Remedies, Prospects." *Brookings Papers on Economic Activity* 1998(1):1-74.

Repullo, Rafael. 2000. "Who Should Act as Lender of Last Resort? An Incomplete Contracts Model." *Journal of Money, Credit and Banking* 32(3 Part 2):580-605.

Rochet, J-C. 1992. "Capital Requirements and the Behaviour of Commercial Banks." *European Economic Review* 36(5):1137-1178.

Shim, Ilhyock. 2006. "Dynamic Prudential Regulation: Is Prompt Corrective Action Optimal? *BIS Working Paper* 206.

Shleifer, Andrei and Robert W. Vishny. 1993. "Corruption." *Quarterly Journal of Economics* 108(3):599-617.

Sleet, Christopher and Bruce D. Smith. 2000. "Deposit Insurance and Lender-of-Last-Resort Functions." *Journal of Money, Credit and Banking* 32: 518-575.

So, Jacky and Jason Wei. 2004. "Deposit Insurance and Forbearance under Moral Hazard." *Journal of Risk and Insurance*

Soral, H. Bartu, Talan B. Iscan and Gregory Hebb. 2006. "Fraud and Banking Crises: Evidence from Micro-Level Transactions Data." *European Journal of Law and Economics* 21:179-197.

Stulz, Rene. 2005. "The Limits of Financial Globalization." *Journal of Finance* 60(4) 1595.

Table 1: *Adding income to the crisis cost regressions*

Equation:	HK2.2		B		C		D		HK3.2		E		F	
Variable	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat
REALINT	0.430	2.8	0.376	2.4	0.401	2.7	0.383	2.8	0.461	2.8	0.454	2.7	0.356	2.7
LIQSUP	0.996	3.3	1.566	3.0	1.324	3.1	0.960	3.2	1.005	2.6	0.984	2.4	0.924	2.4
FORB-A	0.826	2.3	0..483	1.2	0.654	2.6	0.684	2.8	0.882	2.9	0.865	2.9	0.653	2.7
FORB-B	0.994	2.4	1.241	1.4	1.112	3.0	1.151	2.8	0.926	2.0	0.948	2.0	1.200	2.7
GUAR	0.746	2.4	-0.352	0.7					0.923	3.0	0.848	2.0		
LIQSUP*GDPPCPPP			-0.060	1.7	-0.038	1.3								
FORB-A*GDPPCPPP			0.007	0.3										
FORB-B*GDPPCPPP			-0.006	0.1										
GUAR*GDPPCPPP			0.098	3.2	0.074	3.3	0.058	3.4			0.006	0.2	0.042	2.1
Constant	3.426	9.6	3.195	12.4		15.1	3.277	15.9	3.539	10.2	3.516	9.8	3.168	12.6
R-squared / Adj. Rsq	0.589	0.525	0.647	0.533	0.642	0.573	0.628	0.569	0.584	0.520	0.594	0.515	0.617	0.558
S.E.R. / S.S.R.	0.847	22.9	0.839	19.7	0.802	20.0	0.806	20.8	0.851	23.2	0.855	22.7	0.817	21.4
Log likelihood/DW	-44.3	1.87	-41.5	1.79	-41.7	1.83	-42.5	1.95		1.91		1.91		1.90
F-statistic / Prob (F)	9.17	0.000	5.70	0.000	9.28	0.000	10.8	0.000	7.25	0.000	6.00	0.000	7.47	0.000
Method	OLS		OLS		OLS		OLS		2SLS		2SLS		2SLS	

Notes: For the basic data and model see Honohan and Klingebiel (2003).
REALINT is the real interest rate and is included as a macroeconomic control
LIQSUP is where there was open-ended and extensive liquidity support to insolvent institutions
FORB-A and FORB-B are two indicators of forbearabnce: Type A implies that banks were permitted to continue in operation despite open distress.
Type B is a wider category including also eprisodes where other regulations, such as loan loss provisioning were relaxed or not enforced.
GUAR means the government issued an explicit blanket guarantee to depositors or creditors
The sample includes 38 episodes not including Argentina (I) and Egypt.
Dependent variable is log(cost) with mean 1.583 and standard deviation 1.228. The mean of GDPPCPPP is 10.1
All explanatory variables included are shown.
For estimation method 2SLS instruments for LIQSUP and GUAR are: CORRUPT, LAWORDER and (14) dummies for the date on which crises began..
SER=Standard error of regression; SSR=Sum of squared residuals.

Table 2: *Adding survey measures of supervision to the crisis cost regressions*

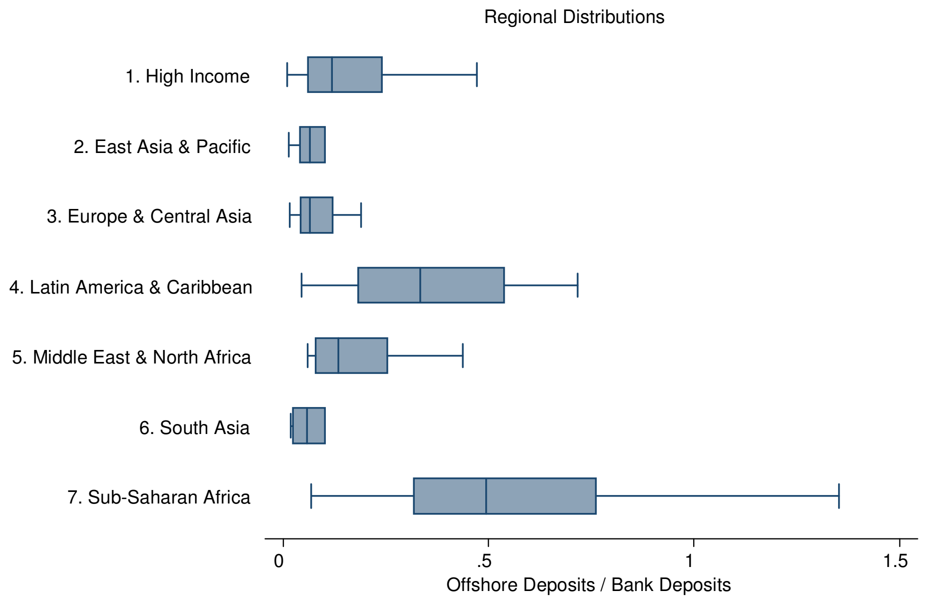
Equation:	G		H		J		K		L		M		N	
Variable	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat	Coeff.	t-Stat
REALINT	0.400	2.3	0.398	2.5	0.392	2.3	0.400	2.1	0.388	2.4	0.410	2.2	0.422	2.5
LIQSUP	0.954	2.8	1.143	3.3	1.150	3.3	1.096	2.6	0.952	2.8	0.864	2.2	0.908	2.5
FORB-A	0.868	2.5	0..881	2.9	0.853	2.5	0.736	1.9	0.744	1.9	0.530	1.4	0.479	1.1
FORB-B	0.976	2.2	0.782	1.7	0.822	1.9	1.075	2.4	0.996	2.3	1.302	2.3	1.243	2.2
GUAR	0.707	2.0	0.617	1.7	0.602	1.7	0.924	2.6	0.652	1.9	0.549	1.3	0.574	1.4
PCPOWER	0.019	0.3			-0.024	0.3	-0.223	1.5						
SFDISCR			0.161	1.2	0.207	1.0	0.589	1.7						
OSPOWER									0.048	0.6			0.043	0.5
PMINDEX											-0.115	1.0	-0.112	0.9
Constant	3.445	7.8	3.212	11.1		5.2	2.022	1.9	2.757	2.6	4.169	9.8	3.668	2.3
R-squared / Adj. Rsq	0.575	0.488	0.585	0.496	0.586	0.478	0.476	0.341	0.573	0.482	0.605	0.492	0.610	0.474
S.E.R. / S.S.R.	0.868	21.8	0.842	19.9	0.857	19.8	0.963	25.1	0.875	23.2	0.515	17.0	0.915	16.7
Log likelihood/DW	-44.3	1.93	-39.8	2.07	-41.7	2.06		1.83	-41.1	1.90		2.61		2.55
F-statistic / Prob (F)	6.55	0.000	6.57	0.000	5.46	0.001	4.36	0.002	6.27	0.000	5.36	0.002	4.47	0.004
Method	OLS		OLS		OLS		2SLS		OLS		OLS		OLS	

Notes: See note to Table 1 for brief explanation of the basic model and data based on Honohan and Klingebiel (2003).

The last four variables are from Barth, Caprio and Levine (2006). PCPOWER is an index of supervisory powers related to prompt corrective action: its sign has been changed so that the higher this index the less intervention must be prompt. SFDISCR is an index of supervisory discretion: the higher this index, the more forbearance. OSPOWER is an index of official supervisory powers. PMINDEX is an index of measures to enhance the effectiveness of private monitoring. These are all measured as of 2001-2.

For estimation method 2SLS instruments for LIQSUP and GUAR are: CORRUPT, LAWORDER and (14) dummies for the date on which crises began..

Figure 1: *Ratio of off-shore to domestic bank deposits*



The figure shows for each region the median, upper and lower quartile and range for the different countries in that region of the ratio of offshore bank deposits held by residents to onshore deposits. The data covers 132 countries and refers to 2004. Source: Calculated from data of *International Financial Statistics* and *Bank for International Settlements* .

Appendix 1: Egypt: Upside-down accounting

It's not usually easy to detect aggressive and misleading accounting practices at banks from simply looking at their accounts. The published accounts of the state-owned banks in Egypt however, displayed a distinctive pattern which suggests a pattern of "upside-down accounting" (de Juan, 2002), where loan-loss provisioning is determined residually in order to achieve a profit target. Specifically, each of the four big state-owned banks which continue to dominate the Egyptian banking system today, though one was privatized in 2006, reported a very small profit each year in the late 1990s and early 2000s despite large loan-loss provisioning. The pattern of their reported profits is only circumstantial evidence but it suggests that loan-loss provisions—probably the element of the accounts that offers the greatest absolute scope for accounting manipulation—was determined by a rule equating it with accounting profit before provisions in order to arrive at a small (close to zero) net profit.

Thus, expressing the sum of profit before tax and loan-loss provisioning for each year as a percentage of year-end total assets for each of the four state banks for which this data was available in the Bankscope database (1994-2001) and plotting this measure of operating profit against loan loss provisioning as a percentage of total assets we find a close linear relationship (Figure 1).[1] A regression analysis gives the following result:

$$\textit{Addition to Loan-loss Provisions} = -0.45 + 0.99\ \textit{Operating Profit}$$
$$(4.5)\quad (15.2)$$
$$RSQ=0.933;\ SEE=0.15.$$

The equation indicates that, whatever the actual procedure for determining loan-loss provisioning, an excellent approximation is obtained by simply subtracting about 0.45 percent of total assets from the operating profit for any given bank in any given year.

In contrast, when we add three private banks for which data is available (from the rating agency Fitch), CIB, Al Watany and EAB they do not fit the pattern at all (Figure 2). For these private banks there is no obvious correlation between operating profit and provisioning.

Despite official statements that loan-loss provisioning at the state banks followed best international practice and adhered to the regulations of the Central Bank of Egypt, this does suggest that the amounts set aside annually into loan-loss reserves were strongly influenced by the availability of operating profits. The not unreasonable conclusion that a much higher level of provisioning was really needed was subsequently confirmed when, after a thorough change of senior management, there were sizable loan write-downs and recapitalization of the banks.

Figure A.1: *Egypt: Operating Profit and Loan-Loss Provisioning at the Four Public Sector Commercial Banks*

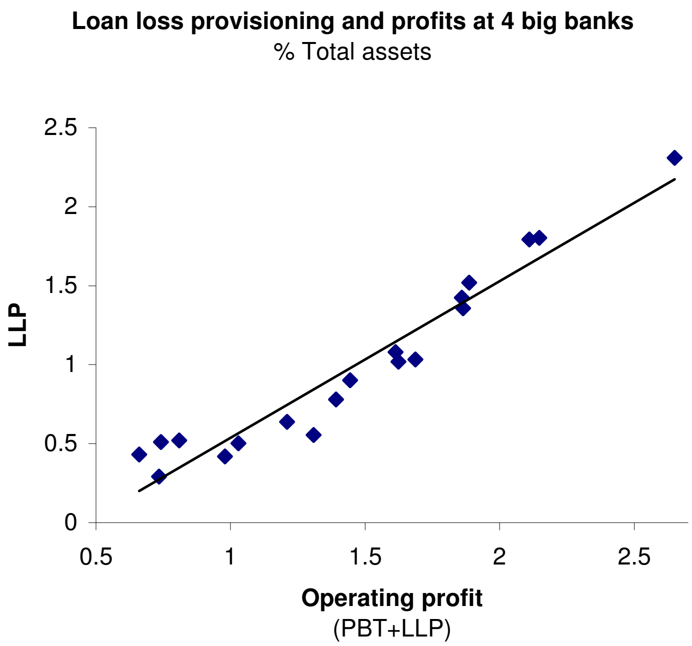
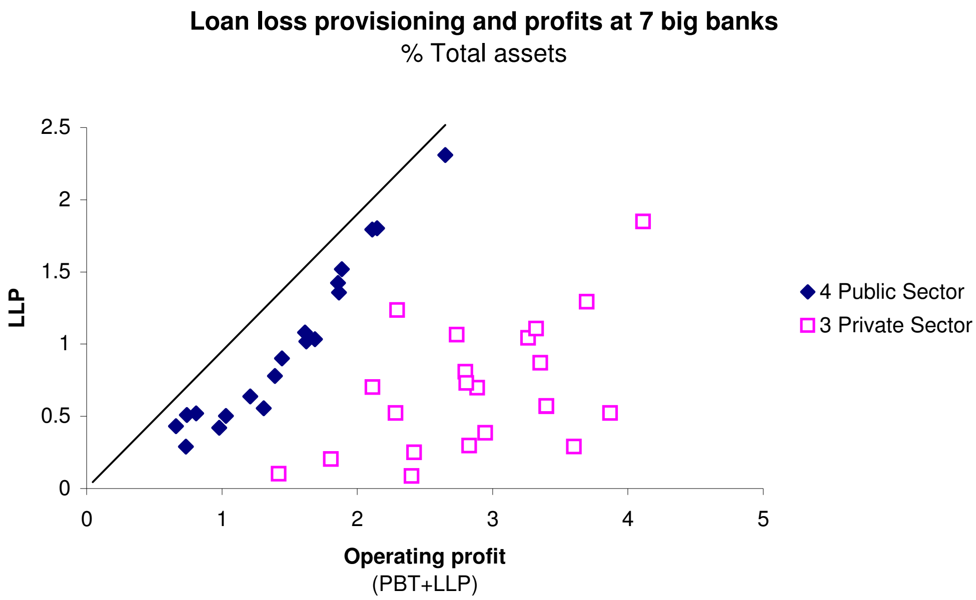


Figure A.2: *Egypt: Operating Profit and Loan-Loss Provisioning at the Four Public Sector Commercial Banks Plus 3 Available Private Banks*



Appendix 2: A model of information, enforcement costs and forbearance

Here is a simple model of regulatory discretion, where regulators seek to minimize the sum of enforcement costs (the pressures placed upon them by the regulated intermediaries trying to avoid the imposition of additional capital requirements) and their exposure to criticism for apparent forbearance in the case of failure.

Imagine a bank operating in discrete time. Each period t results in a new value of the bank's capital. The bank may be required to add capital in an interim period $t+$ before continuing to operate. Suppose that bank insiders, regulators and the general public receive noisy signals about the condition of the bank and that each makes inferences about the capital (net present value of expected future streams) of the bank on the basis of these signals.

To simplify things we will assume right away that bank insiders actually see the true capital position K_t of the bank in period t; their signal is not noisy. However, the regulator does not see this, but observes R_t which we call regulatory capital and which is related to true capital by:

$$R_t = K_t + \varepsilon_t.$$

The fact that the regulator does not see the true capital means that capital may be much lower than the regulator thinks. This can mean much higher risk being taken by the insiders that the regulator expects.

Finally, the general public observes a different indicator P_t which is a noisy indicator of regulatory capital.

$$P_t = R_t + \upsilon_t.$$

The public observes this indicator at all times – both before (t) and after ($t+$) enforcement as described below.

If we take the disturbances ε_t and υ_t to be zero mean Gaussian, the quality of the regulator's information and that of the public can be indexed by their standard deviations, σ_t^{ε} and σ_t^{υ} respectively.

We take it that the stated policy of bank regulation is the maintenance of a statutory minimum capital ratio γ and take the value of γ as predetermined and equivalent to what would be socially optimal in a world without information deficiencies as outlined above (i.e. zero variances σ_t^{ε} and σ_t^{υ}).

We consider two alternative enforcement regimes of capital adequacy. One, the conventional type, in which enforcement is relative to the signal received by the regulator, and "bright line" regulation according to which enforcement is relative to the signal observed by the general public. Like the minimum capital ratio γ, the regime is chosen in advance and cannot be influenced by the regulator. With a conventional enforcement regime, he fact that the public do not see the regulator's signal means that the regulator can forbear without the public being fully aware of this. Given that political or other pressures from the bank insiders mean that requiring recapitalizations represents a costly policy for the regulator, the fact that forbearance can be concealed may influence the action of the regulator. However, with bright line enforcement, forbearance is visible to all.

Enforcement occurs after the signals are observed and takes the form of a required recapitalization, lifting capital to g_t. With a conventional regime, this means that, after enforcement, i.e. at time $t+$, $R_{t+} \geq g_t$. (With bright line enforcement, $P_{t+} \geq g_t$.).

The size of recapitalization is $x_t = Max\{0,\ R_t - g_t\}$ ($x_t = Max\{0,\ P_t - g_t\}$); it is nonzero if and only if $R_t < g_t$ ($P_t < g_t$).

We assume that the insiders have an incentive to minimize capital invested, and will distribute any surplus above required capital. They will then make business decisions which contribute to the evolution of capital between time $t+$ and $t+1$. We assume that

$$K_{t+1} = K_{t+} + u_{t+1}$$

where u_{t+1} is distributed Gaussian with mean zero and standard deviation σ_{t+1}^{u} assumed inversely related to K_t.

We focus on the decisions of the regulator. On the one hand, the regulator does not want to see bank failure, defined as $P_{t+1} < 0$, as this will be associated with social costs and private costs to the extent that the regulator is blamed for forbearance. This we call the regulator's exposure cost. On the other hand, enforcement that involves a nonzero recapitalization requirement x_t will trigger private costs on the regulator as bank insiders use political and other powers to influence the decision. These we will call the regulator's enforcement cost. The regulator is thus seen as choosing $g_t \leq \gamma$ to optimize

$$Y(x_t, \pi_t^F \pi_t^D),$$

where the second term is the product of π_t^F and π_t^D, the probability of failure and the market's inferred probability that the regulator has exercised forbearance (and as such is partly to blame for the failure). For simplicity imagine the cost Y being additive in the enforcement and exposure costs.

The first argument x of the loss function Y is thus the additional capital imposed and its inclusion reflects the pressure on the regulator from the bank shareholders. We could assume that there is a jump in the loss function at $x = 0$ (any forced recapitalization, however minimal, incurs the wrath of the shareholder) and that Y increases thereafter with x. This cost does not depend on either of the information variances. It is illustrated in Figure A2.1.

The probability of failure will depend not only on g but on the uncertainty of the available information σ_t^ε. Thus, the probability of failure, conditional on information available to the regulator at time t+, is

$$\pi_t^F = \text{Prob}\,\{P_{t+1} < 0\}.$$
$$= \text{Prob}\,\{R_{t+1} + v_{t+1} < 0\}.$$
$$= \text{Prob}\,\{K_{t+1} + \varepsilon_{t+1} + v_{t+1} < 0\}.$$
$$= \text{Prob}\,\{K_{t+} + u_{t+1} + \varepsilon_{t+1} + v_{t+1} < 0\}.$$
$$= \text{Prob}\,\{g_t + \varepsilon_t + u_{t+1} + \varepsilon_{t+1} + v_{t+1} < 0\}.$$

The probability of failure thus depends not only on g but also on the variances:

$$\pi_t^F = \text{Prob}\,\{g_t - \varepsilon_t + u_{t+1} + \varepsilon_{t+1} + v_{t+1} < 0 \mid g_t,\ \sigma_{t+1}^u, \sigma_t^\varepsilon, \sigma_{t+1}^\varepsilon\ \ \sigma_{t+1}^v\}.$$

In what follows we assume that the information variances σ_t^ε, σ_t^v are time independent. $\pi_t^F[g_t,\ \sigma_{t+1}^u(g_t - \varepsilon_t), \sigma^\varepsilon, \sigma^v]$

Given that σ_{t+1}^{u} depends on $K_{t+} = g_t - \varepsilon_t$, the probability of failure depends negatively on g_t :

$$\frac{d\pi_t^F\left[g_t, \sigma_{t+1}^u(g_t, \varepsilon_t), \sigma^\varepsilon, \sigma^\upsilon\right]}{dg_t} < 0$$

through two effects: a shift along the normal distribution and a change in the variance. These two effects are illustrated in Figure A2.2, which plots the cumulative distribution function of u_{t+1}.

The dependence of probability of failure on g is plotted in Figure A2.3 for different values of the information variances σ^ε, σ^υ. This shows the dependence on g as convex, as can be deduced from the gaussian distributional assumption. An increase in either variance increases the probability of failure.

Finally, consider the inferred probability that forbearance was exercised, $g<\gamma$. For simplicity we look only at the public's inference from observing P_{t+} ,

$$\text{Prob}\ \{\ g_t<\gamma\ \big|\ P_{t+}\ \}$$

ignoring any additional information deduced from the later observation of P_{t+1}. While the latter is also informative, its information content is degraded by the new shocks u_{t+1}, ε_{t+1}, and υ_{t+1}, whereas $P_{t+} = g_t + \upsilon_{t+}$ so the inference is only complicated by one variance. From this point of view, the simplifying assumption seems acceptable. If the variance σ_t^υ is zero, inference is perfect, as shown in Figure A2.4. (After all, with full information on the part of the public, as with bright line regulation, any value of g below the statutory value makes the public certain that there was forbearance, any value above, makes the public certain that there was no forbearance.) With finite variance σ_t^υ , inferring the enforced capital level $R_{t+} = g$ simply from observation of P_{t+} and the equation $R_{t+} = P_{t+} + v_{t+}$ leads to a Gaussian distribution for R_{t+}, with the probability that forbearance has been exercised $Prob\{R_{t+} < \gamma \mid P_{t+} \} = Prob\{v_{t+} > P_{t+} - \gamma\}$ which selects a point on the cumulative Gaussian distribution of υ_{t+}. If the variance is infinite, the posterior probability distribution of g is very dispersed. Finally, conditional on the actual choice g, the expected value of this probability is an average of the above for values centred around $P_{t+} = g$. Thus, the inferred probability of forbearance is downward sloping in g, but in this case the dependence on σ_t^v is not monotone. These patterns are illustrated in Figure A2.4. A more realistic formulation would build in a Bayesian prior distribution for the public's beliefs about g, but the qualitative patterns would be the same.

The product of the two probabilities is the second argument of the regulator's optimization function, the exposure cost.

Putting together the enforcement and exposure costs, as in Figure A2.5a, shows the two-edged sword of information deficiencies. Even if the regulator imposes the statutory capital requirement, lack of information by the public means that the regulator is still exposed to the risk that they may be blamed for a failure. On the other hand, it is also true that if, faced with severe pressures from the bank shareholders, the regulator avoids imposing the full statutory capital ratio γ, the chances of this forbearance being suspected by the public may be low even in the event of failure. So with low public information and severe pressures from bank shareholders, regulators will tend to impose too little capital. Better information by the public (or almost equivalently, with bright line enforcement) will, in these circumstances, tend to increase the actual capital levels enforced.

To see this, consider the case where public information is weak (high σ_t^v), giving a fairly flat exposure cost, and enforcement costs x are low both as to level and dependence on g. In this case the regulator, driven only by enforcement costs and exposure risk, may even impose a higher than statutory capital requirement for fear that he will be wrongly blamed for forbearance (Figure A2.5b). But with the same public information, a high enforcement cost will tend to lower the enforced level of capital (Figure A2.5c).

With better public information (low σ_t^v), the exposure cost curve is steeper, pushing the enforced capital level g up, even if enforcement costs are high.

Bright line enforcement works in a similar way, driving the enforced capital up whenever the publicly observed indicator P_{t+} lies below γ.

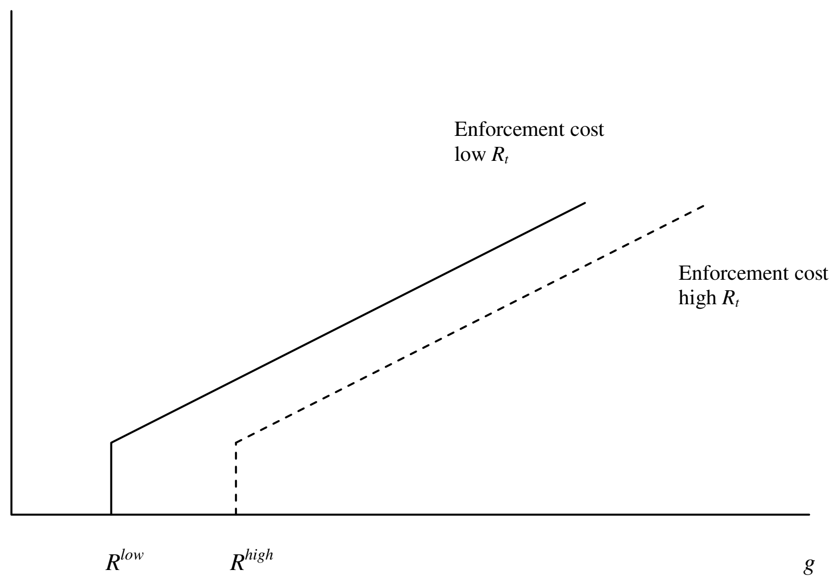


Figure A2.1: *Enforcement cost of requirement* g *at different initial values of capital R*

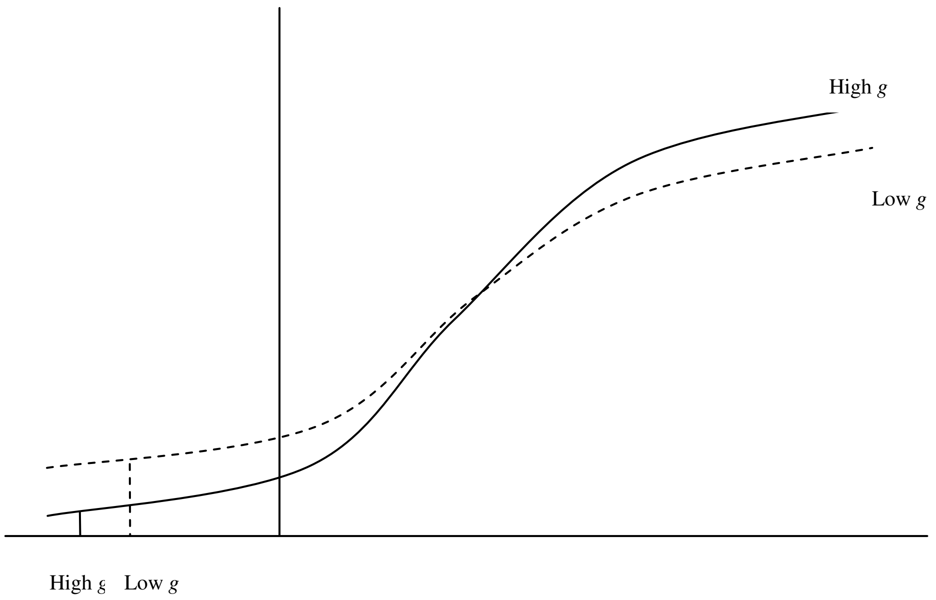


Figure A2.2: *Cumulative distribution of change in capital u depends on enforced capital g*

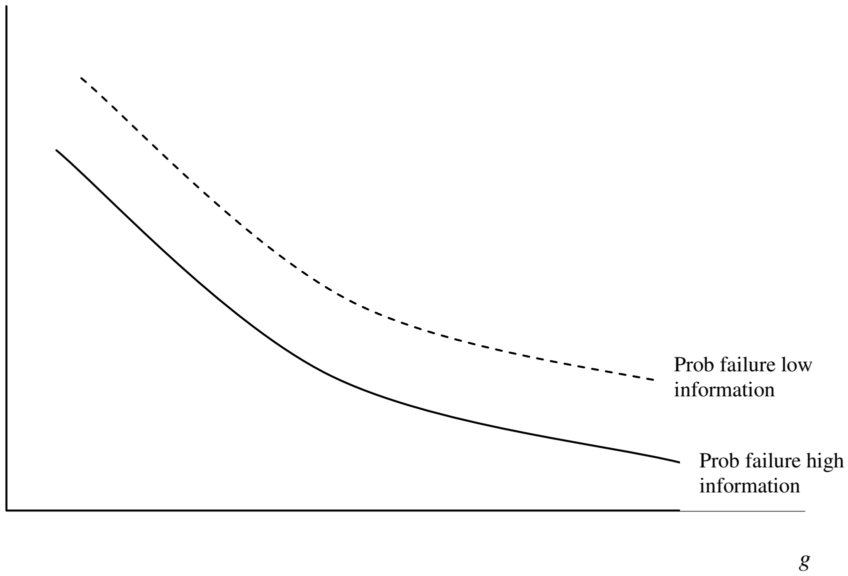


Figure A2.3: *Probability of failure dependence on required capital g*

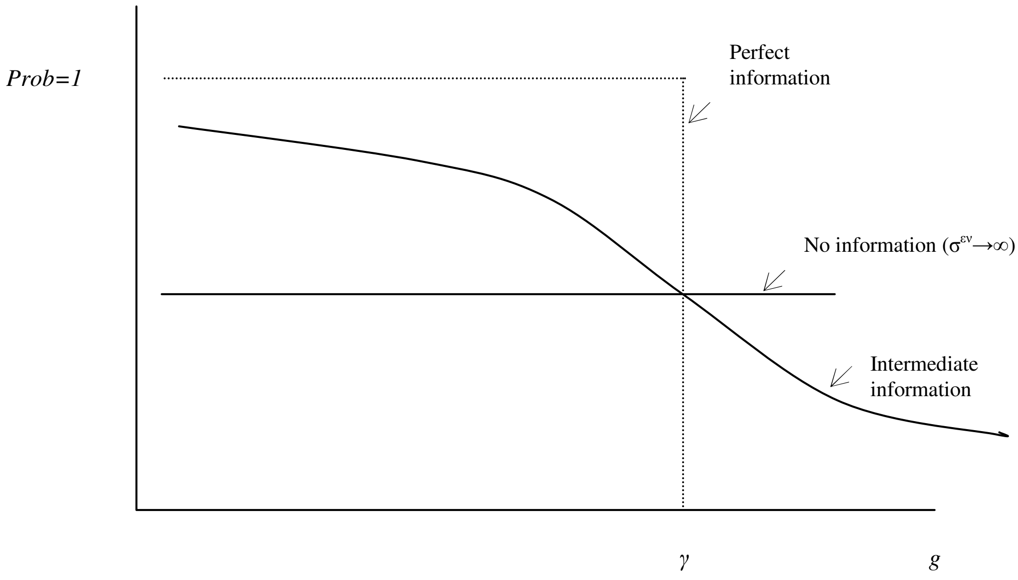
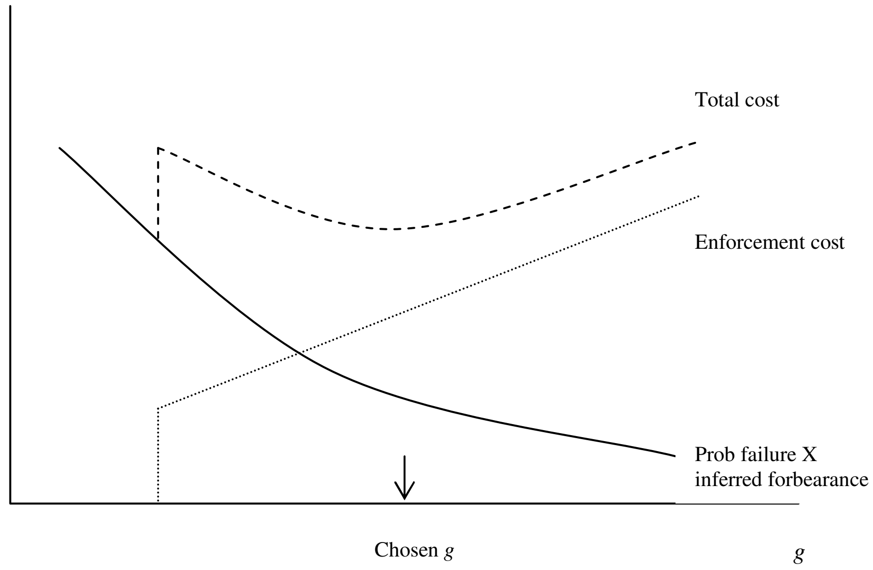
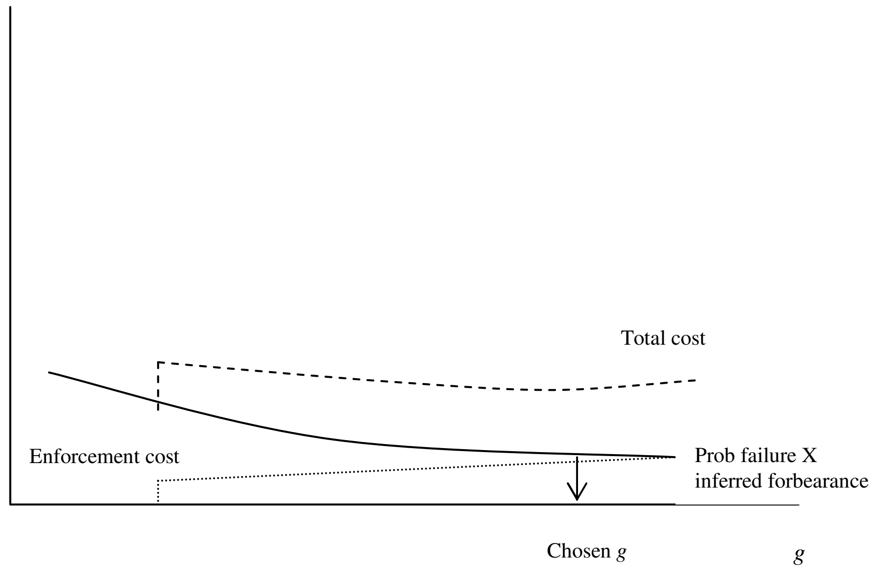


Figure A2.4: *Inferred probability that regulator has exercised forbearance (g<γ)*

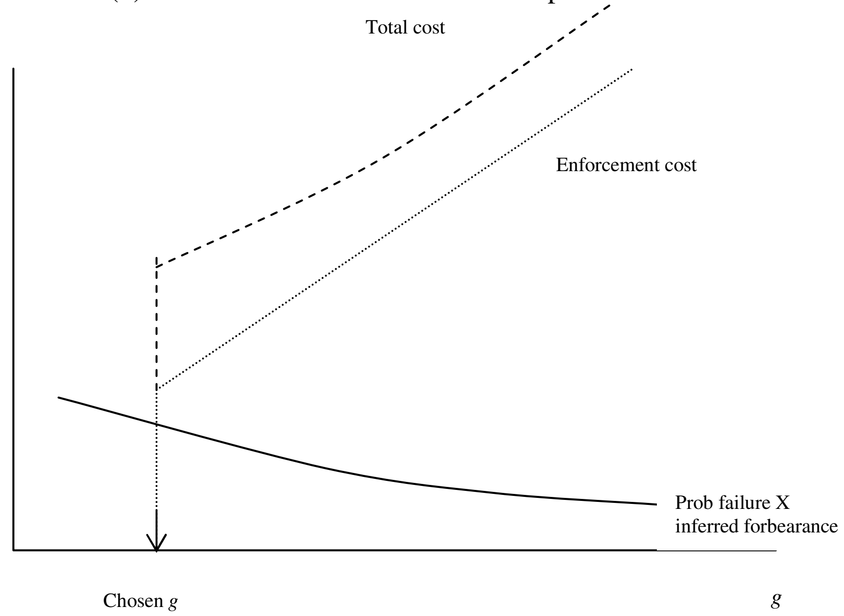
Figure A2.5: *Components of regulator's cost for different levels of enforced capital g*



(a) intermediate values of enforcement and exposure cost



(b) low values of enforcement and exposure cost



(c) High enforcement cost, low exposure cost

Foreign Bank Entry: A Liquidity Based Theory of Entry and Credit Market Segmentation

Nikolaj Schmidt
London School of Economics and Political Science

July 14, 2008

Abstract

This paper analyses how entry by an international bank into a developing economy affects the credit market equilibrium. It offers a novel explanation of how a foreign entrant overcomes asymmetric information problems and complements extant hard vs. soft information based theories of credit market segmentation. In the model, the banks are protected by limited liability. This introduces an agency problem since, in certain states of the world, it is optimal for the banks to lend to negative net present value projects. The agency problem has an asymmetric impact on the local and the foreign bank. The model illustrates how the diversification of the foreign bank's loan portfolio eliminates the agency problem. In contrast, in certain states of the world, the agency problem frustrates the local bank's ability to raise finance. The paper explores the importance of the foreign bank's ability to provide finance during local liquidity shortages, and predicts that foreign entry increases the domestic financial sector's vulnerability to liquidity shocks.

Acknowledgments: The author would like to thank Sudipto Bhattacharya, Amil Dasgupta, Stuart Lewis, Bernardo Guimaraes, Maggie O'Neal, Roselyne Renel and Ashley Taylor as well as the participants in the seminars at the Financial Markets Group at the LSE for their comments and valuable suggestions. The author is grateful for financial support provided under the "Deutsche Bank PhD Fellowship" programme at the London School of Economics and Political Science.

Address: The Financial Markets Group, London School of Economics and Political Science, Houghton Street, London WC2A 2AE, United Kingdom. E-mail: n.f.schmidt@lse.ac.uk

1 Introduction

The merits of allowing foreign banks to operate in developing economies has been a topic of much discussion. Advocates of greater financial integration have argued, that the presence of foreign financial intermediaries lead to better risk management practices and more efficient resource allocations. The opposition has pointed to the financial crisis which followed the financial liberalizations of the 1980's and 90's, and argued that increased competition among financial intermediaries aggravates agency problems and lead to greater financial instability. Currently, India and China are in the process of opening their financial sectors to foreign competition and the countries in Sub-Sahara Africa are experiencing renewed foreign interest in their financial sector. In this paper, I undertake a theoretical analysis of the impact of foreign banks on the financial sector of developing economies. I find that entry by foreign banks can lead to a segmentation of the credit market and that this segmentation, through a reduction of the diversification of the local banks' loan portfolio, can aggravate agency problems and increase financial instability.

Developing economies are characterised by weak institutional infrastructure and opaque reporting standards. This leads to poor information transparency and uncertainty about enforceability of property rights. Given banks ability to alleviate these problems through collateralised lending and borrower screening and monitoring, it is no great surprise that bank lending is the most important source of finance in developing economies. Thus, the analysis of how the credit market and financial stability is affected by the presence of foreign banks is of particular importance in these economies.

This paper is motivated by the observation that the behaviour of foreign banks differ significantly from the behaviour of the local banks. As the local economy goes through a bust, local banks contract credit whilst foreign entrants expand credit.[1] This observation suggests that foreign banks specialize in lending to a segment of the market which remains solvent as the local economy goes through a bust.

This paper aims to address two questions related to how entry by foreign banks affect the local credit market. First, do foreign banks specialise in lending to a particular segment of the local

[1] See for example, Haas and Lelyveld (2006), Peek and Rosengreen (2000), Goldberg et al. (2000).

market? I find that foreign banks, due to the diversification of their business, constitute a more stable source of finance. Their ability to lend as the local economy goes through a bust renders them the financier of choice by local firms with a low exposure to the local business conditions. Second, how do foreign banks affect the stability of the local financial sector? I find that the presence of foreign banks enlarges the set of states in which the local banks are subject to liquidity shortages. In the model, foreign entry leads to a segmentation of the credit market which reduces the diversification of the local banks' loan portfolio. Local banks are subject to limited liability which, in certain states of the world, creates incentives for them to finance unprofitable borrowers. A reduction of the diversification of the local banks' loan portfolio aggravates the agency problem and frustrates the local banks' ability to raise finance. The model explores how the foreign entrant can use the local bank's disability to raise finance to mitigate its inferior information about the local market.

The main idea is as follows. Consider the local economy prior to the entry of the foreign bank. Two types of firms operate in the local economy. One produces for the local market, the other produces for export. The business conditions in the foreign and the domestic market vary independently of each other. To maintain their business, local firms must obtain finance from the local bank which has perfect knowledge about each firm's creditworthiness. The local bank is funded by deposits and operate under limited liability. This introduces the agency problem as it may be optimal for the bank to lend to unprofitable borrowers since it can retain the gains from lending whilst shifting the losses to the depositors. Depositors observe a noisy signal about the creditworthiness of the average local borrower and form rational expectations about the quality of the local bank's loan portfolio. Deposits are not subject to a credible deposit insurance so, if the public signal is sufficiently adverse, it is optimal for the depositors to withdraw their deposits from the bank. If the firms that produce for the local market constitute the majority of the firms in the economy, the public signal tends to be unfavourable when the business conditions in the local market are poor. This is so even if the business conditions are prosperous for firms producing for export. Thus, in specific states of the world, the local bank fails to raise deposits and creditworthy firms which produce for export are denied finance.

The foreign entrant is subject to limited liability and is funded by deposits. The entrant has less information about the local economy than the incumbent, and its lending decisions can be made contingent only on the borrowers type and on the public signal. The foreign bank is active in many different economies and therefore has a well diversified portfolio. This reduces the agency problem arising from limited liability and ties down the depositors' expectations about the quality of the foreign bank's portfolio.[2] When the diversification is sufficiently large, the agency problem disappears and the foreign entrant can always raise deposits. As the local economy goes through a bust, the public signal turns negative and local depositors withdraw from the local bank. In these states, the foreign bank becomes the monopoly lender to solvent local firms. The prospect of extracting monopoly rents in future periods allows the foreign entrant to contest the incumbents information advantage. Due to the fraction of firms producing for the local market, the public signal is highly correlated with the state of the local economy. Solvent local borrowers' disability to obtain finance from the local bank is therefore more pronounced for firms producing for export. Consequently, the foreign bank's prospect of extracting monopoly rents is higher when it finances exporters than when it lends to firms producing for the local market. If the expected monopoly profits from lending to exporters is sufficient for the entrant to overcome the incumbents information advantage, and if the expected monopoly profits from lending to firms producing for the local economy are such that the incumbent's information advantage prevails, the credit market becomes segmented. The foreign bank lends to firms which produce for export and the local bank lends to firms which produce for the local market.

A number of theoretical studies analyse the occurrence of clientele effects in credit markets with heterogeneous banks. These studies suggest that, when banks are heterogeneous with respect to size, the credit market becomes segmented along the lines of hard and soft information.[3] Large

[2] The impact of diversification has similarities with the mechanism explored in Diamond (1984) and Cerasi and Daltung (2000). In Diamond (1984), the financial intermediary is efficient since the diversification of its portfolio eliminates a non pecuniary cost. As in Cerasi and Daltung (2000), in the model presented below diversification reduces an agency problem within the financial intermediary.

[3] Hard information can be easily quantified and distributed to third parties. Typically, hard information can be found in a firm's annual report. Soft information cannot be quantified or transmitted to third parties. This type of information includes

banks lend to borrowers which produce hard information and small banks lend to borrowers which supply soft information. When applied to the analysis of foreign bank entry, the large bank has been interpreted as the foreign entrant. In Stein (2002), local loan officers need the managements' approval to grant loans. The incentives to collect information about borrowers are impaired when there is a high likelihood that the information is soft and therefore can not be transmitted to the bank's management. In small banks where the loan office is himself part of the management, the incentives to gather soft information are restored. In Berger and Udell (2002), local loan officers are subject to an agency problem which prevents them from lending based on soft information. This agency problem increases with the distance between the banks' management and the local loan officer. Williamson (1988) suggests that the optimal organizational structure is different for a bank that specialises in lending to hard information borrowers relative to a bank which specialises in lending to soft information borrowers.

In the model presented in this paper, the banks are homogeneous with respect to their ability to process information. The segmentation of the credit market arises since the diversification of the foreign bank's business provides it with a more stable source of funds, which can be used to finance local firms in states where the local financial system is subject to liquidity shortages. Thus, the theory presented in this paper complements extant theories and argues that loans extended by the foreign entrant exhibit features which are considered value added by a specific segment of the credit market.

The driving force behind the credit market segmentation has similarities with the theory put forward in Froot, Sharfstein and Stein (1993). Froot et al. analyse a firm's optimal risk management decision and conclude that risks which are adversely correlated with the firm's future business opportunities should be hedged. In the model presented below, firms producing for export borrow from the foreign entrant since this isolates them from liquidity shortages in states where their business opportunities are good.

This paper also contributes to the discussion about how competition between financial interme-

whether the loan applicant appears trustworthy and hard working.

diaries affect financial stability. In the model, the market segmentation reduces the diversification of the local bank's portfolio. This aggravates the agency problem and increases the local bank's exposure to liquidity shortages. The change in the diversification illustrates a novel mechanism through which competition between financial intermediaries amplify agency problems and increases financial instability. Allen and Gale (2004) finds that increased competition between financial intermediaries reinforces incentives for risk shifting. In Boot and Thakor (2000), Boot and Marinč (2007), and Allen, Carletti and Marquez (2008), the increased competition between financial intermediaries reduces the banks' return on costly screening and monitoring effort and increases financial instability.[4]

Finally, the model is related to the literature on competition between financial intermediaries under asymmetric information.[5] The analysis is cast in an asymmetric information framework and has features which resemble Dell'Ariccia and Marquez (2004) and Hauswald and Marquez (2006). In these models, the entrant uses either a lower cost of funds or an exogenously specified ability to provide some value added service to contest the incumbent's information advantage.[6] In the model presented below, the entrant and the incumbent pay the same risk adjusted price for funds and the value added service delivered by the entrant, the ability to provide finance during a bust, arises endogenously. Similar to Sengupta (2007) and Bester (1985), I assume that the borrowers have access to collateral which can be used as a costly sorting mechanism.

The paper is organized as follows. Section 2 presents the benchmark model. Section 3 derives the equilibrium prior to the entry of the foreign bank. Section 4 analyses the foreign entrant's behaviour and characterises the equilibrium with foreign bank entry. Section 5 outlines the theory's empirical implications and reviews the empirical evidence. Section 6 concludes. Proofs are relegated to the appendix.

---

[4] See Allen and Gale (2004) for a literature review.

[5] See Broecker (1990) for early contributions to this literature.

[6] In Hauswald and Marquez (2006), banks are located on the circumference of a unit circle. The distance between the bank and the borrowers is an inverse measure of the information the bank can obtain about the borrower. The bank's choice of location is interpreted not only in a physical sense, but also in the sense of location in the product space.

2 The benchmark model

2.1 Basic setup

Consider an economy with three time points ($t = 0, 1, 2$), a continuum of firms ("borrowers"), one local bank and a set of depositors. All agents are risk neutral and borrowers are segmented into two types. Type D borrowers, with measure λ, produce goods for the domestic market and type I borrowers, with measure $(1 - \lambda)$, produce goods for export. I assume that the measure of type D borrowers exceeds the measure of type I borrowers, i.e. $\lambda > \frac{1}{2}$.

By the beginning of each period, each type of borrower has access to one investment project. The projects have a tenure of one period and require one unit of finance. Firms are cashless and consume the returns from the projects immediately, so to initiate a new project each firm must obtain a loan from the bank. I assume that the firms have a stock of productive assets which can be pledged as collateral.[7] There are two potential outcomes of the projects, success and failure, and two different qualities, good and bad. Good projects succeed with probability P_G and bad projects succeed with probability P_B, where $P_G > P_B$. Successful projects return X and unsuccessful projects return 0. Only good projects are creditworthy, i.e. $P_G X - R_f > 0$ and $P_B X - R_f < 0$, where R_f is the rate of return on a risk free alternative investment. The quality of the projects varies across time, such that a type i borrower, $i \in \{D, I\}$, with a good project at time t may hold a bad project at time $t + 1$. I assume that the variation is independent over time and that the probability of obtaining a good project is equal across types, i.e. $\Pr(P_D = P_j) = \Pr(P_I = P_j)$ for $j \in \{G, B\}$, where P_D is the quality of a type D borrowers project and P_I is the quality of a type I borrower's project.[8] All borrowers of a given type hold the same project.

At time t, the depositors can either place money on deposit with the bank or invest in the risk free alternative. I assume that there is no deposit guarantee scheme and that deposits is the only source of finance to the bank.

If the projects that the bank finances in the first period fails, the bank continues to operate in

[7] The assets are assumed pivotal to the investment project and can not be sold to finance the investment.

[8] The assumption of independent identically distributed projects eases the exposition of the main idea. The models qualitative conclusions are robust to assuming a positive (but not perfect) correlation both between the projects and across time.

the second period. The bank's discount factor is normalised to one.

2.2 Information structure

The borrowers' types are observable to all agents and all borrowers know the quality of their project. If the bank is the most recent lender to a particular borrower (the "relationship lender"), it can observe the quality of the borrower's project. I assume that the local bank can observe the quality of all the borrowers' projects at time 0.

All agents observe a noisy signal about credit quality of the average borrower's project. The signal, γ_t, is given by

$$\gamma_t = \lambda P_{Dt} + (1-\lambda) P_{It} + \tilde{\varepsilon}_t,$$

where $P_{Dt}, P_{It} \in \{P_G, P_B\}$ and $\tilde{\varepsilon}_t$ is a noise term,

$$\tilde{\varepsilon} \sim N\left(0, \sigma_{\varepsilon}^2\right).$$

All information is revealed simultaneously at the end of each period such that γ_t and P_t is known just prior to time t.

Depositors, and banks without prior lending relationships, observe only the public signal, γ_t.[9]

2.3 Financial contracts

The bank offers one period loan contracts to the borrowers. The borrowers hold productive assets so lending can be subject to collateral requirements. Collateral is subject to a liquidation inefficiency which implies that the bank upon liquidation of the collateral can capture only a fraction $\Delta < 1$ of the face value of the collateral.[10] A loan contract is given by a tuple (R_t, C_t) where R_t and C_t is respectively the interest rate and the collateral requirement on a loan at time t.[11] It is costless for the bank to verify both the outcome of the project and whether the borrower has diverted

[9] In the benchmark model, there is only one local bank which is the relationship lender to all borrowers. The possibility of a non-relationship lender arises following the entry of the foreign bank.

[10] One rationale for the collateral inefficiency is, that the current owner of the assets has the best knowledge about the redeployability of the asests. Alternatively, the inefficiecy arises because the assets are most productive with the current owner and therefore can only be sold at a discount.

[11] All interest rates are gross rates, i.e. they include both return of principal and accrued interest.

funds from the project. Loan contracts are subject to limited recourse, so the bank can only secure repayment from the returns on the project and by liquidating the collateral.

The demand deposits offered at time t carry interest $\delta(\gamma_t)$ and can be redeemed at time $t+1$.

2.4 Discussion of model setup

In the model, the firms producing for the local market and the firms producing for export are subject to different shocks. The business conditions in the local economy can differ from the business conditions in the export markets. This motivates that both the quality and the outcomes of the projects vary as a function of whether the firm produces for export or for local consumption.

The lack of deposit insurance in the local economy goes against empirical observations, but it reflects that the deposit insurance may only cover partial losses or that settlement from the deposit insurance may be subject to severe delays. Demirgüç-kunt et al. (2005) find that all banks are subject to either explicit or implicit deposit insurance, but that the deposit insurance most often is only partial and subject settlement delays. The run on Northern Rock in September 2007 underscores that, when depositors have a low degree of confidence in the quality of the deposit insurance they may behave as it is absent.

The information structure illustrates that the banks learn about the borrowers' business through lending. Bank loans are subject to covenants which gives the lender access to non-public information about the borrowers' business. The additional information obtained through relationship lending is modelled via the assumption that only the relationship lender observes the quality of an individual borrower's project.

If the projects financed by the bank in the first period fail, the bank continues its operations in the second period. This assumption is motivated by to aspects of weak institutional infrastructure. First, given the information structure, it is ex post efficient for a financial regulator to allow the local bank to continue its operations in the second period. A weak institutional infrastructure may obstruct the financial regulator's ability to commit to an ex ante policy rule. Second, if the repercussions against a loan officer that approved loans which failed in the first period are low, say if it is easy for the loan officer to get a position at another financial institution, the loan officer may behave as if the bank continued to operate in period two.

3 Credit market equilibrium without foreign banks

The equilibrium consists of a set of optimal actions and rational beliefs for each of the agents. Prior to characterizing the equilibrium, I impose a constraint which uniquely determines the banks financing decision. Let the state variable P_t be given by $P_t \equiv (P_{Dt}, P_{It})$.

Lemma 1 *Under the parameter constraints, $R_f > \lambda x$, $P_G > \lambda$ and $1 - P_B > \lambda > P_B$, the local bank lends only to type D borrowers when $P_t = (P_B, P_B)$ and only to borrowers with good projects when $P_t \in \{(P_B, P_G), (P_G, P_B), (P_G, P_G)\}$.*

I impose the parameter constraints in Lemma 1. By fixing the banks financing decision, these parameter constraints eases exposition of the main idea.

3.1 The bank's problem

There are no information imperfections in the benchmark model so, since collateral liquidation is inefficient, it is optimal for the bank to sort the borrowers without the use of collateral. There is only one bank in the benchmark model, so borrowers obtain credit at the monopoly lending rate, $R_t = X$. The bank's optimal lending behaviour follow directly from Lemma 1. To maximize profits, the bank offers demand deposit at a rate which render depositors indifferent between financing the bank and investing in the risk free alternative.

3.2 The depositors' problem

Prior to the decision of whether to invest in demand deposits or in the risk free alternative, the depositors observe the signal γ_t and form beliefs about the quality of the borrowers projects. Given the realisation of the public signal, the depositors beliefs are given by,

$$\Pr\left(P_t = (P_j, P_k)\,|\gamma_t\right) = \frac{\phi\left(\frac{\gamma_t - \lambda P_j - (1-\lambda)P_k}{\sigma_\varepsilon}\right)\Pr\left(P_D = P_j\right)\Pr\left(P_I = P_k\right)}{\Pr\left(\gamma_t\right)}, \tag{1}$$

where $j, k \in \{G, B\}$, $\phi(\cdot)$ is the partial distribution function of the standard normal distribution, and

$$\Pr(\gamma_t) = \sum_{j,k \in \{G,B\}} \phi\left(\frac{\gamma_t - \lambda P_j - (1-\lambda) P_k}{\sigma_\varepsilon}\right) \Pr(P_D = P_j) \Pr(P_I = P_k)$$

is the probability of observing γ_t.

For all values of the public signal, depositors assign a positive probability to the occurrence of state $P_t = (P_B, P_B)$. In this state, it is optimal for the local bank to lend to borrowers with bad projects and shift the expected losses to the depositors. The bad projects are unprofitable, so this lending strategy leads to an expected loss on the bank's loan portfolio. Consequently, when depositors become sufficiently certain that state (P_B, P_B) has materialised, i.e. when the public signal is sufficiently adverse, it is optimal for the depositors to withdraw from the bank. Thus, there is a threshold value of the public signal, γ^*, such that the local bank fails to raise demand deposits if $\gamma_t < \gamma^*$. γ^* is the lowest value of the public signal such that depositors are indifferent between investing in demand deposits and in the risk free alternative. The expected return on demand deposits is increasing in $\delta(\gamma_t)$, so, since $\delta(\gamma_t) \leq X$, γ^* solves[12]

$$[\Pr(P = (P_B, P_B) | \gamma_t) P_B + (1 - \Pr(P = (P_B, P_B) | \gamma_t)) P_G] X = R_f. \tag{2}$$

It is straight forward to verify that there is a unique value for γ^*.[13] For $\gamma \geq \gamma^*$, depositors invests in demand deposits if $\delta \geq \delta^*(\gamma_t)$, where $\delta^*(\gamma_t)$ solves,

$$\begin{aligned}[][\Pr(P_t = (P_B, P_G) | \gamma_t) + \Pr(P_t = (P_G, P_B) | \gamma_t)] P_G \delta_t \\ + \Pr(P_t = (P_G, P_G) | \gamma_t) \left[P_G^2 \delta_t + P_G (1 - P_G) X\right] + \Pr(P_t = (P_B, P_B) | \gamma_t) P_B \delta_t = R_f.\end{aligned}$$

Given the public signal and the bank's lending decision, $\delta^*(\gamma_t)$ is the return on demand deposits which ensures that depositors are indifferent between investing in demand deposits and in the risk

[12] Note that for $P_t \neq (P_B, P_B)$, the expected return to depositors is given by,

$$P_G^2 X + P_G (1 - P_G) \lambda X + P_G (1 - P_G)(1 - \lambda) X = P_G X.$$

[13] The existence of a unique solution to 2 follows since, the left hand side goes to $P_B X < R_f$ for $\gamma = -\infty$ and to $P_G X > R_f$ for $\gamma \to \infty$. Since the left hand side is continous and strictly increasing in γ, there is a unique solution for γ.

free alternative. The first term on the left hand side of this expression measures the expected return from investing in demand deposits when only one type of the borrower holds a good project.[14] The second term measure the expected return when both types of borrowers hold good projects. The constraints imposed on λ implies that depositors are repaid in full only if both types of borrowers succeed.[15] The third term on the left hand side measures the expected return to depositors from states where both types of borrowers hold bad projects.

3.3 The borrowers' problem

Firms are protected by limited liability and therefore always apply for a loan. The bank can verify whether the borrower diverts money from the project, so the borrower invests in the project when the loan application is successful.

Proposition 2, characterises the equilibrium of the benchmark model.

Proposition 2 *In equilibrium, depositors' beliefs are given by* (1) *and depositors invest in demand deposits if and only if* $\gamma_t \geq \gamma^*$ *and* $\delta_t \geq \delta^*(\gamma_t)$. *Subject to the availability of deposits, the bank lends to borrowers with good projects if* $P_t \in \{(P_G, P_B), (P_B, P_G), (P_G, P_G)\}$ *and only to type* D *borrowers if* $P_t = (P_B, P_B)$. *The bank lends at the monopoly rate,* $R_t = X$, *and offers demand deposits at* $\delta_t = \delta^*(\gamma_t)$. *Borrowers always apply for a loan and invest in the project if the loan application is successful.*

The bank rejects loan applications from borrowers with good projects only when the public signal causes depositors to withdraw from the bank. The inefficient financing decision, i.e. the event that the borrowers hold good projects but fail to be financed, is a consequence of the depositors' attempts to ensure that the bank lends only to projects with a positive net present value. Since $\Pr(P_D = P_G \cap \gamma < \gamma^*) < \Pr(P_I = P_G \cap \gamma < \gamma^*)$, type I borrowers are more frequently exposed to

[14] Reacall that by Lemma 1, the bank finances only borrowers with good projects if $P_t \in \{(P_G, P_B), (P_B, P_G)\}$.

[15] Recall that $R_f > \lambda X > (1 - \lambda) X$.

inefficient termination of finance than type D borrowers. To see this, note that

$$\begin{aligned}\Pr\left(P_D = P_G \cap \gamma < \gamma^*\right) &= \Pr\left(P_D = P_G\right)\Pr\left(\gamma < \gamma^* | P_D = P_G\right) \\ &= \Pr\left(P_D = P_G\right)\left[\Pr\left(P_I = P_G\right)\Phi\left(\frac{\gamma^* - P_G}{\sigma_\varepsilon}\right)\right. \\ &\left. + \Pr\left(P_I = P_B\right)\Phi\left(\frac{\gamma^* - \lambda P_G - (1-\lambda) P_B}{\sigma_\varepsilon}\right)\right],\end{aligned}$$

and,

$$\begin{aligned}\Pr\left(P_I = P_G \cap \gamma < \gamma^*\right) &= \Pr\left(P_I = P_G\right)\Pr\left(\gamma < \gamma^* | P_I = P_G\right) \\ &= \Pr\left(P_I = P_G\right)\left[\Pr\left(P_D = P_G\right)\Phi\left(\frac{\gamma^* - P_G}{\sigma_\varepsilon}\right)\right. \\ &\left. + \Pr\left(P_D = P_B\right)\Phi\left(\frac{\gamma^* - \lambda P_B - (1-\lambda) P_G}{\sigma_\varepsilon}\right)\right],\end{aligned}$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Since $\lambda > \frac{1}{2}$, the public signal is more adverse in state (P_B, P_G) that in state (P_G, P_B). Consequently, depositors are more likely to withdraw from the bank when type I borrowers hold good project than when type D borrowers hold good projects. Thus, inefficient termination is more likely for type I borrowers than for type D borrowers.

The inefficient termination of finance creates demand for a more stable source of finance.

4 Credit market equilibrium and foreign bank entry

This section analyses how entry by a foreign bank affects the credit market equilibrium.

4.1 The foreign bank

The foreign bank has a well established international presence and can bid for deposits and allocate capital to projects in a range of economies. I assume that the local bank's ability to raise deposits is unaffected by the presence of the foreign bank.

The foreign bank is active, i.e. raises deposits and extends credit, in N economies which are isomorph to the benchmark economy. I assume that the realisation of the state variable, P_t, is independent across the economies.[16] The information structure implies that the foreign bank, upon entry, can make its lending decision contingent on the borrowers type and on the public signal. As the local bank, the foreign bank can build relationships with the local borrowers which allows it to observe the quality of the borrowers' projects. I assume that the foreign bank has been active in the other N economies for at least one period. Further, I assume that the depositors in the N economies know the number of markets in which the foreign bank is active, and that they observe the public signal of their own economy only. Under these assumptions, the optimal behaviour of the foreign bank and of the depositors in the N economies can be characterised.

Proposition 3 *If* $\Pr(P_I = P_G) > 0$*, then there exists a number of economies,* N^**, such that the foreign bank can always raise deposits when it is active in more than* N^* *economies.*

The logic behind Proposition 3 relies on three lemmata. First, if the bank finances bad projects, it will do so in a limited number of economies only. It is profitable for the bank to finance bad projects because it can shift the losses to the depositors. The outcomes of the projects are independent across the economies, so if the bank finances bad projects in many economies, the law of large numbers drives its expected profits to zero. In contrast, by financing bad projects in a small number of economies, the bank can prevent the law of large numbers from kicking in and can obtain positive expected profits. Thus, since it is optimal for the bank to lend to bad borrowers in a limited number of economies only, its expected profits from financing borrowers with bad projects are limited. Second, when $\Pr(P_I = P_G) > 0$, the banks expected profits from financing only good projects is strictly increasing in N. Consequently, there is a value of N such that the expected profits from financing only good projects exceeds the expected profits from financing only bad projects. By similar logic, there is a value of N such that the profits from financing only good projects exceeds the profits from financing both good and bad projects. Thus, when N is sufficiently large, the bank finances only borrowers with good projects. Last, when N is large, rational depositors correctly anticipates that the bank refrains from financing bad projects and therefore the foreign bank can

[16] The independence assumption is made for expositional purposes only. The model's conclusions require only that the state of the economies is not perfectly correlated.

always raise deposits. In effect, the depositors disregard their public signal as they know that the bank has incentives to finance good projects only.

I assume that $N > N^*$, such that the foreign bank never fails to raise deposits. Let δ_I^* be the deposit rate offered by the foreign bank. The foreign bank finances only borrowers with good projects and, by the law of large numbers, the realised return on its portfolio equals the expected return so,

$$\delta_I^* = R_f.$$

4.2 Equilibrium under foreign bank entry

The model is solved by backwards induction. When the banks compete for borrowers, the sequential structure is as follows. At stage one, the banks simultaneously submit a bid for the borrower. The outcome of the first round bidding is observable to both lenders. At stage two the relationship lender is granted the opportunity to improve on its first round bid. This sequential structure ensures the existence of a pure strategy equilibrium.

Prior to the analysis of the banks' problem, Lemma 4 lists the contracts through which the banks can sort the borrowers. Recall that a loan contract is a tuple (R, C), where R is the lending rate and C is the collateral requirement.

Lemma 4 *There exists a contract, the "optimal collateral contract", which is accepted only by borrowers with good projects and which maximizes the return to borrowers with good projects. Under this contract, the lender's expected profits are zero. If a bank offers the optimal collateral contract, and the competitor observes the quality of the borrowers' projects, then the competitor can offer a contract,* $\left(\tilde{R}, 0\right)$, *which is weakly preferred by all borrowers with good projects, and which yields an expected profit of* $E(\pi^c)$, *where* $E(\pi^c) > 0$. *Further, there exists a contract, the "monopoly collateral contract", which is accepted only by borrowers with good projects and under which the lender expects to obtain the monopoly profits.*

The appendix lists the contracts of Lemma 4 explicitly. The higher failure probability of borrowers with bad projects implies that it is more costly them to post collateral. Consequently, by adjusting the collateral requirements, the bank can design a contract which is accepted only by borrowers with good projects. Due to the collateral inefficiency, $\Delta < 1$, the cost of extending a collateralised loan to a borrower with a good project exceeds the break-even cost of financing a borrower with a good project. Thus, a bank which observes the quality of the borrowers projects can offer a non-collateralised loan to borrowers with good projects and obtain positive expected profits.

4.3 The banks' problem under foreign bank entry

The outcome of the game at $t = 0$ implies that the banks, at $t = 1$, compete under asymmetric information. The relationship lender observes the quality of the borrowers' projects. This gives it an information advantage since the non-relationship lender must use the public signal and form Bayesian beliefs about the quality of the borrowers' projects. At $t = 1$, if the non-relationship lender offers an uncollateralised contract, then the relationship lender matches this contract if it is profitable and refrains from financing the borrower if the contract is unprofitable. Thus, to protect itself against adverse selection, the non-relationship lender offers the optimal collateral contract and obtains zero expected profits. The relationship lender offers the contract $\left(\tilde{R}, 0\right)$ and obtains an expected profit of $E\left(\pi^c\right)$.

At $t = 0$, the local bank is the relationship lender to both types of borrowers. The entrant can offer either the optimal collateral contract and obtain zero expected profits or a non-collateralised, type contingent contract. Under the latter (the "pooling contract"), the entrant finances all borrowers of a given type without knowing whether they hold good or bad projects. To attract borrowers with good projects, the terms of the pooling contract must be advantageous relative to the terms of the contract offered by the relationship lender. However, since the entrant can only make Bayesian inference about the quality of the borrower's project, on average it finances some measure of borrowers with bad projects. Under the pooling contract, all borrowers are financed on the same terms, so the contract is unprofitable to the entrant in the first period. The benefit of the pooling contract is, that the entrant becomes the relationship lender in the subsequent period. It is advantageous for the entrant to offer the pooling contract if this leads to positive expected profits across the two periods of the game.

In a rational expectations equilibrium, the contracts offered by the banks affect the depositors' behaviour. To solve the model, I conjecture that, in equilibrium, the credit market is segmented. Subsequently, I provide conditions under which this conjecture holds true.

Conjecture 5 *In equilibrium, the foreign bank lends to the type I borrowers and the local bank lends to the type D borrowers.*

Let γ^{**} be the threshold value of the public signal below which depositors withdraw from the local bank, and let $R^j_{t,k}$ and $\pi^j_{t,k}$ be respectively the lending rate and the expected profits to bank k from lending to type j borrowers in period t. Let π^j_k be the present value of bank k's profit from lending to a type j borrower, i.e. $\pi^j_k \equiv \Sigma_t \pi^j_{t,k}$. The most competitive non-collateralized contract bank I will offer to a type j borrower , $j \in \{D, I\}$, yields zero expected profits, so

$$\begin{aligned} \pi^j_I &= \left(\bar{P}_{0,j} R^j_{1I} - R_f\right) + E\left(\pi^j_{2,I}\right) = 0 \iff \\ R^j_{1I} &= \frac{R_f}{\bar{P}_{0,j}} - \frac{1}{\bar{P}_{0,j}} E\left(\pi^j_{2I}\right), \end{aligned}$$

where,

$$E\left(\pi^I_{2,I}\right) = \Pr\left(P_I = P_G \cap \gamma_1 \geq \gamma^{**}\right) E\left(\pi^c\right) + \Pr\left(P_I = P_G \cap \gamma_1 < \gamma^{**}\right)\left(P_G X - R_f\right),$$

$$E\left(\pi^D_{2,I}\right) = \Pr\left(P_D = P_G \cap \gamma_1 \geq \gamma^{**}\right) E\left(\pi^c\right) + \Pr\left(P_D = P_G \cap \gamma_1 < \gamma^{**}\right)\left(P_G X - R_f\right),$$

and

$$\bar{P}_{0,j} = \Pr\left(P_j = P_G | \gamma_0\right) P_G + \Pr\left(P_j = P_B | \gamma_0\right) P_B.$$

The most competitive contract the D-bank will offer a type j borrower with a good project yields zero expected profits, so

$$\begin{aligned} \pi^j_D &= P_G\left(R^j_{1D} - \delta_D\left(\gamma_0\right)\right) + E\left(\pi^j_{2D}\right) = 0 \iff \\ R^j_{1D} &= \delta_D\left(\gamma_0\right) - \frac{1}{P_G} E\left(\pi^j_{2D}\right), \end{aligned}$$

where,

$$E\left(\pi_{2,D}^{I}\right)=\Pr\left(P_I=P_G\cap\gamma_1\geq\gamma^{**}\right)E\left(\pi^c\right),$$

$$E\left(\pi_{2,D}^{D}\right)=\Pr\left(P_D=P_G\cap\gamma_1\geq\gamma^{**}\right)E\left(\pi^c\right)$$
$$+\Pr\left(P_D=P_B\cap\gamma_1\geq\gamma^{**}\right)P_B\left(X-E_0\left[\delta_1|\gamma_1\geq\gamma^{**},P_D=P_B\right]\right).$$

The difference between the two banks' most competitive non-collateralised contracts is driven by two effects. First, under the pooling contract, on average the entrant finances some measure of borrowers with bad projects. This raises the default rate on the entrant's portfolio and thereby its breakeven lending rate. This effect is measured by the term $\frac{1}{\bar{P}_{0,j}}>\frac{1}{P_G}$. Second, both banks draw on their expected future profits to improve their lending rate at time zero. The foreign bank's expected future profits are given by the relationship lender's information rents, $E\left(\pi^c\right)$, plus the monopoly profits from states where the local bank fails to raise deposits. The local banks expected future profits is given by the information rents plus the expected profits from states where local depositors erroneously supply deposits to the local bank.

Conjecture 5 is true when $R_{1D}^{D}\geq R_{1I}^{D}$ and $R_{1I}^{I}\leq R_{1D}^{I}$.

Lemma 6 *The entrant finances type I borrowers and the incumbent finances type D borrowers if,*

$$\delta_D\left(\gamma_0\right)-\frac{R_f}{\bar{P}_{0,D}}-\left(\frac{1}{P_G}-\frac{1}{\bar{P}_{0,D}}\right)\Pr\left(P_D=P_G\cap\gamma_1\geq\gamma^{**}\right)E\left(\pi^c\right)$$
$$+\frac{\Pr\left(P_D=P_G\cap\gamma<\gamma^{**}\right)\left(P_GX-R_f\right)}{\bar{P}_{0,D}}$$
$$-\frac{\Pr\left(P_D=P_B\cap\gamma\geq\gamma^{**}\right)P_B\left(X-E_0\left[\delta_1|\gamma_1\geq\gamma^{**},P_D=P_B\right]\right)}{P_G}<0 \quad (3)$$

and

$$\frac{R_f}{\bar{P}_{0,I}}-\delta_D\left(\gamma_0\right)-\left(\frac{1}{\bar{P}_{0,I}}-\frac{1}{P_G}\right)\Pr\left(P_I=P_G\cap\gamma\geq\gamma^{**}\right)E\left(\pi^c\right)$$
$$-\frac{\Pr\left(P_I=P_G\cap\gamma<\gamma^{**}\right)\left(P_GX-R_f\right)}{\bar{P}_{0,I}}\leq 0 \quad (4)$$

Condition (3) and (4) follow from the requirements $R_{1D}^{D}\geq R_{1I}^{D}$ and $R_{1I}^{I}\leq R_{1D}^{I}$. Intuitively, when the foreign bank finances type I borrowers, its expected profits from states where it is a monopoly lender exceeds its expected losses from the pooling contract. Conversely, when the local bank lends to type D borrowers, the foreign bank's losses from the pooling contract exceeds its expected profits from states where it is the monopoly lender.

The local bank is more frequently forced to reject loan applications from type I borrowers with good projects than from type D borrowers with good projects. Therefore, the foreign bank's expected monopoly profits from type I borrowers exceeds its expected monopoly profits from type D borrowers. The losses from the pooling contract are independent of the borrowers type, so it is more profitable for the foreign bank to extend the pooling contract to type I borrowers than to type D borrowers. Therefore, if the entrant poaches type D borrowers, it also poaches type I borrowers. The converse however, is not true. In the following I assume that (3) and (4) are fulfilled, so that both banks remain active in equilibrium.

If the local bank fails to raise deposits in the second period, the foreign bank can capture the monopoly rents irrespective of the contracts it offered in the first period. Consider the strategy where the foreign bank offers the optimal collateral contract in the first period and the monopoly collateral contract in the second period. Under this strategy, it avoids the loss from the pooling contract and obtains the monopoly profits in states where the local bank fails to raise deposits. This strategy does not lead to a segmentation of the credit market and the local bank therefore fails to attract deposits for $\gamma < \gamma^*$. In contrast, if the entrant offers the pooling contract at time $t = 0$, the credit markets become segmented and the local bank fails to raise deposits for $\gamma < \gamma^{**}$. Thus, if $\gamma^{**} > \gamma^*$ and if the rents from being the monopoly lender are sufficiently high, it is optimal for the entrant to offer the pooling contract at time $t = 0$. Lemma 7 lists the condition such that it is optimal for the entrant to offer the pooling contract.

Lemma 7 *The entrant offers the pooling contract at* $t = 0$ *if and only if*

$$\left(1 - \frac{\bar{P}}{P_G}\right)\left(\Pr\left(P_I = P_G \cap \gamma_2 \geq \gamma^{**}\right)\pi^c - R_f\right) + \left(P_G X - R_f\right)\left[\Pr\left(\gamma_2 < \gamma^{**} \cap P_I = P_G\right) - \Pr\left(\gamma_2 < \gamma^* \cap P_I = P_G\right)\right] \geq 0. \quad (5)$$

Condition (5) ensures that the increase in the expected monopoly profits from offering the pooling contract at time $t = 0$ (the second term in (5)) exceeds the losses incurred under the pooling contract (the first term in (5)). The first term is negative, so a necessary condition for condition (5) to be fulfilled is $\gamma^* < \gamma^{**}$. That indeed this is the case is verified in Lemma 8.

To maximize profits, the local bank offers demand deposits at a the rate which render depositors indifferent between demand deposits and the risk free alternative. The local bank lends only to type D borrowers so the equilibrium deposit rate, $\delta^{**}(\gamma_t)$, solves

$$[\Pr(P_D = P_B|\gamma_t) P_B + \Pr(P_D = P_G|\gamma_t) P_G] \delta_t = R_f.$$

4.3.1 The depositors' problem

A necessary condition for the depositors to finance the local bank is, that the risk adjusted return on demand deposits weakly exceeds the return on the risk free alternative. The highest deposit rate that the local bank can credibly offer depositors is given by $\delta(\gamma_t) = \tilde{R}$, so the threshold value of the public signal, below which the local bank fails to raise deposits, γ^{**}, solves

$$[\Pr(P_D = P_B|\gamma_t) P_B + (1 - \Pr(P_D = P_B|\gamma_t)) P_G] \tilde{R} = R_f. \qquad (6)$$

This equation can be solved numerically for γ^{**}, but is suffices to note that it yields a unique solution for γ^{**}.[17] The difference between γ^* and γ^{**} is driven by two effects. First, the foreign bank's presence forces the local bank to offer a more competitive lending rate. This leads to a reduction in the deposit rate that the local bank can credibly promise depositors, i.e. the maximum deposit rate decreases from X to $\tilde{R}$.[18] Second, prior to the foreign bank's entry, the local bank lends to bad borrowers only in state $P_t = (P_B, P_B)$. The foreign bank lends to type I borrowers, so post foreign bank entry, it is optimal for the local bank to lend to borrowers with bad projects in the states $P_t \in \{(P_B, P_G), (P_B, P_B)\}$. The aggravation of the local bank's agency problem reduces

[17] To see the existence of a unique solution for γ^{**}, note that for $\gamma \to \infty$, the left hand side of the equation goes to $P_G\tilde{R} > R_f$. For $\gamma \to -\infty$, the left hand side of the equation goes to $P_B\tilde{R} < R_f$. Thus the existence of a unique solution follows from the observation that the left hand side is continous and strictly increasing in γ.

[18] Although the local bank can lend to borrowers with bad projects at the monopoly rate, it can only promise a deposit rate of $\tilde{R}$. If it promises a deposit rate in excess of $\tilde{R}$, the local bank reveals that it is lending to borrowers with bad projects, which will cause depositors to withdraw from the bank.

the depositors' willingness to finance the bank. Both of these effects tend to increase the threshold value of the public signal.

Lemma 8 *Foreign bank entry frustrates the local bank's ability to raise deposits, i.e.* $\gamma^* < \gamma^{**}$.

Lemma 8 completes the characterisation of the credit market equilibrium under foreign bank entry.

Proposition 9 *Under condition* (3), (4) *and* (5), *the foreign bank offers the pooling contract to type I borrowers and the optimal collateral contract to type D borrowers. Foreign entry leads to a segmentation of the credit market. The foreign bank finances type I borrowers at rate* R_D^I *and the the local bank finances type D borrowers at rate* $\tilde{R}$. *Entry by the foreign bank frustrates the local bank's ability to raise deposits, i.e.* $\gamma^{**} > \gamma^*$.

The equilibrium reveals a feature which is novel to the literature on competition between financial intermediaries under asymmetric information. In extant models, the entrant contests the incumbent's information advantage through cheaper funding, better screening technology or through the ability to deliver an exogenously specified value added service.[19] In the model presented here, the entrant and the incumbent pays the same risk adjusted price for deposits and have the same ability to learn about the borrowers' credit quality. The entrant's ability to finance creditworthy borrowers in states where the local financial system is subject to liquidity shortages permits it to mitigate the incumbents information advantage. In particular, foreign finance is important to local firms with a low exposure to the local business conditions. The models predictions are primarily relevant for banks in developing economies, since these are more frequently subject to liquidity shortage.[20]

The logic behind Lemma 8 suggests that competition is more severe following entry by a foreign financial intermediary than following entry by a local financial intermediary. The foreign bank

[19] See for example the one period models presented in Sengupta (2006) or Dell'Ariccia and Marquez (2004) for an analysis of the importance of funding levels, Claeys and Hainz (2006) for an analysis based on differences in screening technology and Hauswald and Marquez (2006) for an analysis based on the ability to deliver a value added service.

[20] Through 1980-2005, the average standard deviation of real cost of deposits was 1.6% in G7 countries but 12.9% in 25 major emerging markets. The standard deviation of real demand deposit growth was 14% for G7 economies and 24% for the 25 major emerging markets (See Mian and Khwaja (2006) for further discussion). For a discussion of the correlation between the variance in deposits and variance in bank credit see Mian and Khwaja (2006) or Berlin and Mester (1999).

increases competition via two channels. First, the foreign bank offers local borrowers an alternative source of finance which forces the local bank to give up some of its monopoly rents. Second, foreign entry leads to a segmentation of the credit market which frustrates the local bank's ability to raise deposits. Entry by a local financial intermediary affects competition only via the first channel. In particular, since the foreign bank finances type I borrowers, competition for type I borrowers increases more following entry by a foreign bank than following entry by a local bank.

4.4 Welfare considerations

The foreign bank enhances the efficiency of the local financial system by eliminating the set of states where firms with good projects fail to raise finance. On the flip side, the local bank's information about type I borrowers is discarded and, on average, firms with bad projects are financed under the foreign bank's pooling contract. Overall, welfare is improved if the failure to finance good projects is costly relative to the cost of financing bad projects. That is, welfare is improved if

$$\Pr(\gamma < \gamma^*)\left[\lambda \Pr(P_D = P_G | \gamma < \gamma^*) + (1-\lambda)\Pr(P_I = P_G | \gamma < \gamma^*)\right](P_G X - R_f) \\ - (1-\lambda)\Pr(P_I = P_B)(R_f - P_B X) \geq 0$$

The stylized framework presented in this paper ignores many costs and benefits of foreign bank entry. As illustrated in the model, foreign entry can lead to a contraction of the lending spreads.[21] In the model the demand for credit is inelastic, but if profitable projects are heterogeneous, lower lending spreads can increase welfare by enlarging the measure of entrepreneurs with good projects that apply for credit.[22]

Extant literature illustrates that competition among financial intermediaries can lead to greater financial instability.[23] In the model, the foreign bank's presence frustrates the local bank's ability to raise deposits. In states where the local bank is subject to liquidity shortages, the foreign bank finances all good projects. Thus, a higher frequency of local liquidity shortages has no welfare

[21] The lending spread is defined as the difference between lending rates and deposit rates.

[22] For empirical support, see Demirgüç-Kunt et. al. (1999), Claessens et. al. (2001) or Peria et. al.(2002)

[23] See Allen and Gale (2004) for a survey.

effects. These observations imply that the foreign bank's ability to expand credit during local liquidity shortages is insufficient to conclude that foreign entry enhances the stability of the local financial system. As the model highlights, the foreign entrant may itself indirectly cause the liquidity shortage. Further, the foreign bank's behaviour during the liquidity shortage have large redistributional effects which are generally not neutral from a welfare perspective.[24] BIS (2004) and BIS (2005) argue, that the presence of a reputable foreign financial institution may enhance the stability of the local economy by allowing local depositors' flight to quality without negative impact on the capital account. Proposition 3 provides a theoretical rationale for this line of reasoning.

5 Empirical implications and evidence

5.1 Empirical implications

The model has a set of testable implications. First, under the credit market segmentation, firms that produce for export obtain credit from the foreign entrant. If firms that produce for export are larger than the firms that produce for local consumption, as is frequently the case, this prediction coincides with the prediction from the information based theories of credit market segmentation. Second, the foreign banks' supply of finance is more stable than the local banks supply of finance, and it does not contract as local business conditions deteriorate. Third, following entry by a foreign bank, local banks are more likely to fail to raise finance. Last, foreign banks improve the financing conditions more for local firms producing for export than for local firms producing for the local market. Consequently, the market value of firms that produces for export should respond more positively to the entry of a foreign bank than the market values of the firms which produce for local consumption.

5.2 Empirical evidence

Due to data constraints, the empirical tests of clientele effects following the entry of foreign banks into developing markets are rather limited and rests primarily on results from surveys and interviews

[24] The model predicts that the foreign entrant makes substantial rents during the bust which reduces local firms realized profits.

with bank managers. One exception is Mian (2006) which analyses a detailed data set for Pakistan. Mian (2006) finds, that local banks lend to small firms and foreign banks lend to large firms. This is interpreted as evidence that information fictions and agency problems prevents the foreign entrant from lending to small firms. The dataset however, also indicates that sectors with large exports tend to borrow from foreign banks and therefore, it does not reject the hypothesis presented in this paper.

Based on questionnaire surveys and interviews with bank managers, Galac and Kraft (2000) finds that one of the most important activities of foreign entrants in Croatia was import-export financing. In addition, some of the foreign bank managers stated that they financed only exporters. Konopielko (1999) conducts a survey among foreign bank managers in Poland, Hungary and Czech Republic and finds that the foreign banks' main objectives were to finance foreign trade and support existing clients.[25] Haas and Naaborg (2005a) surveys managers of foreign banks with a presence in Central Europe and the Baltic States and finds that, upon entry, the foreign banks objective was to finance multinational firms. In addition, their surveys indicate that, as a result of increasing competition, the foreign banks' objective changed over time.

A wide range of empirical literature analyses how the presence of foreign banks affect the amplitude of the host economies' business cycles. Overall, this literature finds that foreign banks continue to lend as the local economy goes through a recession.

Haas and Naaborg (2005a) and Haas and Lelyveld (2006) find, that foreign banks in Central and Eastern Europe maintained credit outstanding during the financial turmoil in the 1990s. Galac and Kraft (2000) finds that foreign banks with a physical presence in Croatia expanded both their direct lending and their supply of liquidity in the interbank market during the banking crisis of the late 1990s. Consistent with the model's predictions, foreign entrants in Croatia appear to have made considerable profits from their operations during the banking crisis.

Goldberg et al. (2000) analyse data for Mexico and Argentina and find that, during the periods of financial unrest in the 1990s, the credit growth of foreign banks was less volatile than the credit

[25] As a specific example, Citibank was the 20th bank in terms of loan volumes but the second largest bank in terms of foreign trade finance (Konopielko (1999)).

growth of local banks. These periods of unrest were characterised by depositor flight to quality, and the authors interpret their findings as evidence that access to a more stable source of funds allowed the foreign banks to maintain their credit outstanding. Peria and Moody (2002) and Peek and Rosengren (2000) support this finding in their analysis of a range of Latin American countries. Goldberg (2001) analyses the lending behaviour of US banks in emerging markets and find this to be uncorrelated with the real demand cycles of the local markets. The author interprets this as evidence that US banks with a physical presence in emerging markets tend to maintain their supply of credit when the local economy goes through a bust.

6 Conclusion

Extant theories predict that a bank's size and organizational structure creates agency problems which can lead to clientele effect. Large banks are predicted to finance borrowers which produce hard information and small banks are predicted to finance borrowers which produce only soft information. When applied to the analysis of foreign bank entry, the large bank has been interpreted as the entrant. This paper has presented an alternative theory of the clientele effects following foreign bank entry. The theory emphasises that distinct features of the foreign bank's business renders it well suited to finance local borrowers with a high exposure to international business conditions. The diversification of the foreign bank's business provides it with a stable source of finance and permits it to maintain credit outstanding as the local economy goes through a bust. This creates a segmentation of the credit market, as the ability to raise finance during a downturn of the local economy is important to firms whose business opportunities have a low correlation with the state of the local economy. Thus, foreign banks finance firms with a low exposure to the local business conditions and local banks finance firms with a high exposure to the local business conditions.

The segmentation is along risk factors, so it reduces the diversification of the local bank's loan portfolio. This aggravates the local bank's agency problems and adds to its vulnerability to liquidity shortages.

The models empirical implications find some support in existing empirical work with the caveat that much of the evidence on market segmentation, due to data constraints, is based on survey data.

To assess the welfare implications of foreign competition, it is important to understand how the presence of foreign banks affect the local credit market. This paper is silent on a range of questions which are important in this respect. For example, how important is the foreign bank's mode of entry (greenfield versus acquisition versus cross border lending)? And, how does the entrant and the incumbent interact in the deposit and interbank market? This paper has illustrated that there is a fundamental difference between the local bank and the foreign entrant. Thus, increased competition due to entry by a foreign bank is different from increased competition due to an increase in the number of local players. An analysis of these questions can shed more light on how foreign banks affect the local financial system and on the particular features of the local economy which may render foreign bank entry welfare enhancing.

7 Appendix

Proof. Lemma 1. The parameter constraint, $R_f > \lambda X > (1-\lambda) X$ implies, that when the bank finances both types of borrowers, they must both succeed for the bank to repay depositors in full. The bank is protected by limited liability, so in state (P_B, P_B), the expected return from financing only type D borrowers is

$$\lambda P_B (X - \delta).$$

$\lambda > \frac{1}{2}$, so $\lambda P_B (X-\delta) > (1-\lambda) P_B (X-\delta)$ and therefore, if the local bank finances only one type of borrower, it finances the type D borrowers. Under limited liability, financing both types of borrowers yields a return of

$$P_B^2 (X - \delta),$$

and since $\lambda > P_B$, it follows that

$$\lambda P_B (X - \delta) > P_B^2 (X - \delta).$$

Thus, in state (P_B, P_B) the local bank finances only type D borrowers.

In state (P_G, P_B), the return to financing only type D borrowers is given by

$$\lambda P_G (X - \delta).$$

The return to financing type I borrowers is

$$(1-\lambda) P_B (X - \delta),$$

and since $\lambda > (1-\lambda)$ and $P_G > P_B$, it follows that if the bank finances only one type of borrower, it finances type D borrowers. Under limited liability, the return from financing all borrowers is given by

$$P_B P_G (X - \delta).$$

Since $\lambda > P_B$,

$$\lambda P_G (X - \delta) > P_B P_G (X - \delta),$$

so the local bank finances only type D borrowers.

In state (P_B, P_G), the return to financing type D borrowers is given by

$$\lambda P_B (X - \delta),$$

and the return to financing only type I borrowers is given by

$$(1 - \lambda) P_G (X - \delta).$$

Since $(1 - \lambda) > P_B$ and $P_G > \lambda$ it follows that, if the local bank finances only one type of borrower, it finances type I borrowers. The return from financing both types of borrowers is,

$$P_G P_B (X - \delta),$$

and since $(1 - \lambda) > P_B$ it follows that

$$(1 - \lambda) P_G (X - P_G) > P_G P_B (X - \delta).$$

Thus, in state (P_B, P_G), the local bank only finances type I borrowers.

Last, in state (P_G, P_G) the return to the limited liability bank from financing both types of borrowers is given by

$$P_G^2 (X - \delta),$$

and the return from financing only type D borrowers is,

$$\lambda P_G (X - \delta),$$

so since $P_G > \lambda$, it follows that

$$P_G^2 (X - \delta) > \lambda P_B (X - \delta).$$

Thus, in state (P_G, P_G) the local bank finances both types of borrowers (it follows straight forward that it is never optimal for the local bank to finance type I borrowers only). This verifies the claim in the lemma. ■

Proof. Proposition 2. The proof follows directly from Lemma 1 and the presiding discussion. ■

Proof. Proposition 3. The proof consists of three lemmata which in conjunction verify the proposition.

Lemma 10 *If the foreign bank finances loan applicants with bad projects, it will do so in a finite number of economies only.*

Proof. Assume that all borrowers in all economies hold bad projects. Let n be the number of economies in which the foreign bank lends, and let i be the number of economies with successful outcomes. Let $E(\pi_t(n))$ be foreign bank's expected profits. The bank's profits are increasing in the lending rate, so assume that the bank lends at X. Assume that the foreign bank finances only type I borrowers (that indeed this is the case is verified in 9). Limited liability implies that the foreign bank obtains positive profits if $i > \frac{\delta_I}{X}n$, where δ_I is the foreign banks deposit rate. Thus, the optimal n solves,

$$\max_n E(\pi_t(n)) = \max_n (1-\lambda) \sum_{i=\frac{\delta_I}{X}n}^{n} \binom{n}{i} P_B^i (1-P_B)^{n-i} (iX - n\delta_I).$$

It follows that,

$$\begin{aligned}
&(1-\lambda) \sum_{i=\frac{\delta_I}{X}n}^{n} \binom{n}{i} P_B^i (1-P_B)^{n-i} (iX - n\delta_I) \\
&\qquad = (1-\lambda) n \sum_{i=\frac{\delta_I}{X}n}^{n} \binom{n}{i} P_B^i (1-P_B)^{n-i} \frac{(iX - n\delta_I)}{n} \\
&\qquad \leq (1-\lambda)(X-\delta_I) n \sum_{i=\frac{\delta_I}{X}n}^{n} \binom{n}{i} P_B^i (1-P_B)^{n-i} \\
&\qquad = (1-\lambda)(X-\delta_I) n \Pr\left(i \geq \frac{\delta_I}{X}n\right) \\
&\qquad = (1-\lambda)(X-\delta_I) n \Pr\left(\frac{i}{n}X \geq \delta_I\right).
\end{aligned}$$

$\lim_{n\to\infty} \frac{i}{n} = P_B$, so $\lim_{n\to\infty} \frac{i}{n}X = P_B X < R_f < \delta_I$. Thus, a $\upsilon > 0$ can be found such that $(P_B + \upsilon) X < \delta_I$, and, for every such υ, there exists an $n < \infty$, $\tilde{n}$, such that $\left|\frac{i}{\tilde{n}} - P_B\right| < \upsilon$. Consequently, $\Pr\left(\frac{i}{\tilde{n}}X \geq \delta_I\right) \leq \Pr((P_B+\upsilon) X \geq \delta_I) = 0$ so $n \Pr\left(\frac{i}{n}X \geq \delta_I\right) = 0$ for all $n \geq \tilde{n}$ and therefore $E(\pi_t(n)) = 0$ for $n \geq \tilde{n}$. $E(\pi_t(n))$ is closed and $E(\pi_t(1)) = (1-\lambda) P_B (X - \delta_I) > 0$, so there is an optimal n, n^*, with $n^* \in [1, \tilde{n}[$. ■

Lemma 11 *Let α be the average fraction of borrowers with good projects. If $\alpha > 0$, then there exists a value N^{**} such that the foreign bank lends only to borrowers with good projects if $N > N^{**}$.*

Proof. Let β be the average fraction of type I borrowers with good projects in each economy, i.e. $\beta = \alpha(1-\lambda)$, and β^c be the average fraction of type I borrowers with bad projects in each economy, i.e. $\beta^c = (1-\alpha)(1-\lambda)$. Let $E(\pi_t|(y_G, y_B))$ be the foreign bank's expected profits when it finances good projects with a measure y_G and bad projects with a measure y_B. Let i be the number of economies where type I borrowers with good projects are successful and l be the number of economies where type I borrowers with bad projects are successful. The bank's profits from lending to borrowers with good projects are positive when $i > \frac{\delta_I}{X}\alpha N$. By the law of large numbers, the number of economies where type I borrowers hold good projects is αN. Thus since the foreign bank is protected by limited liability,

$$E(\pi_t|(\beta N, 0)) = \sum_{i=\frac{\delta_I}{X}\alpha N}^{\alpha N} \binom{\alpha N}{i} P_G^i (1-P_G)^{\alpha N - i} (i(1-\lambda)X - \beta N \delta_I)$$

Let n^* be number of economies in which the foreign bank finances borrowers with bad projects. By the previous lemma, n^* is finite. Then,

$$E(\pi_t|(\beta N, \beta^c n^*))$$

$$\begin{aligned} = \max[& \sum_{l=0}^{\frac{\delta_I}{X}(1-\alpha)n^*} \binom{n^*}{l} P_B^l (1-P_B)^{n^*-l} (l(1-\lambda)X - \beta^c n^* \delta_I) \\ & + \sum_{l=\frac{\delta_I}{X}(1-\alpha)n^*+1}^{n^*} \binom{n^*}{l} P_B^l (1-P_B)^{n^*-l} (l(1-\lambda)X - \beta^c n^* \delta_I) \\ & + \sum_{i=0}^{\frac{\delta_I}{X}\alpha N} \binom{N}{i} P_G^i (1-P_G)^{N-i} (i(1-\lambda)X - \beta N \delta_I) \\ & + \sum_{i=\frac{\delta_I}{X}\alpha N+1}^{\alpha N} \binom{\alpha N}{i} P_G^i (1-P_G)^{\alpha N-i} (i(1-\lambda)X - \beta N \delta_I), 0] \end{aligned} \quad (7)$$

Note that the first two terms is the return to financing borrowers with bad projects in n^* economies. Since n^* is finite, the value of these two expressions is finite. The third and fourth terms is the return on financing borrowers with good projects in αN economies. Thus,

$$\sum_{i=0}^{\frac{\delta_I}{X}\alpha N^{**}} \binom{N^{**}}{i} P_G^i (1-P_G)^{N^{**}-i} (i(1-\lambda)X - \beta N^{**}\delta_I) + \sum_{i=\frac{\delta_I}{X}\alpha N^{**}+1}^{\alpha N^{**}} \binom{\alpha N^{**}}{i} P_G^i (1-P_G)^{\alpha N^{**}-i} (i(1-\lambda)X - \beta N^{**}\delta_I)$$
$$= (1-\lambda) X P_G \alpha N^{**} - \beta N^{**}\delta_I$$
$$= (1-\lambda)\alpha N^{**}[P_G X - \delta_I].$$

Since n^* is finite and the outcome of the projects is independent across the economies, there is a value of N, N^{**}, such that $\delta_I \leq P_G X$. Therefore, the expected return from financing borrowers with good projects is increasing in N^{**}, so there must be a value of N^{**} such that the third and the fourth term in (7) exceeds the first and the second term in (7). Consequently, when $N \geq N^{**}$, the foreign bank's limited liability can be ignored. Further, the third term in (7) goes to zero as N^{**} increases. To see this, note that

$$\sum_{i=0}^{\frac{\delta_I}{X}\alpha N} \binom{\alpha N}{i} P_G^i (1-P_G)^{\alpha N-i} (i(1-\lambda)X - \beta N\delta_I) \leq 0,$$

and

$$\begin{aligned} &\sum_{i=0}^{\frac{\delta_I}{X}\alpha N} \binom{\alpha N}{i} P_G^i (1-P_G)^{\alpha N-i} (i(1-\lambda)X - \beta N\delta_I) \\ \geq\ & -\delta_I \beta N \sum_{i=0}^{\frac{\delta_I}{X}\alpha N} \binom{\alpha N}{i} P_G^i (1-P_G)^{\alpha N-i} \\ =\ & -\delta_I \beta N \Pr\left(i \leq \alpha N \frac{\delta_I}{X}\right) \\ =\ & -\delta_I \beta N \Pr\left(\frac{i}{\alpha N} \leq \frac{\delta_I}{X}\right) \end{aligned}$$

By the law of large numbers, $\lim_{N\to\infty} \frac{i}{\alpha N} = P_G$, so for any $\upsilon \leq P_G(1-P_G)$, there is a value of N, N^{**}, such that $\frac{i}{\alpha N^{**}} \geq P_G - \upsilon$. For $N \geq N^{**}$,

$$-\delta_I \beta N \Pr\left(\frac{i}{\alpha N} \leq \frac{\delta_I}{X}\right) \leq -\delta_I \beta N \Pr\left(P_G - \upsilon \leq \frac{\delta_I}{X}\right).$$

As noted, since n^* is finite, the diversification of the foreign bank's portfolio implies that $\lim_{N^{**}\to\infty} \delta_I \leq \frac{R_f}{P_G}$. Thus,

$$\begin{aligned}
\lim_{N^{**}\to\infty} \Pr\left(P_G - \upsilon \leq \frac{\delta_I}{X}\right) &\leq \Pr\left(P_G - \upsilon \leq \frac{\frac{R_f}{P_G}}{X}\right) \\
&\leq \Pr\left(X\left(1 - \frac{P_G(1-P_G)}{P_G}\right) \leq R_f\right) \\
&= \Pr\left(P_G X \leq R_f\right) = 0
\end{aligned}$$

Therefore,

$$-\delta_I \beta N^{**} \Pr\left(P_G - \upsilon \leq \frac{\delta_I}{X}\right) = 0,$$

and the third term of (7) goes to zero as claimed. Consequently,

$$\begin{aligned}
E\left(\pi_t | \left(\beta N, \beta^c n^*\right)\right) &= E\left(\pi_t | \left(\beta N, 0\right)\right) \\
&\quad + \sum_{i=0}^{\frac{\delta_I}{X}(1-\alpha)n^*} \binom{n^*}{i} P_B^i \left(1-P_B\right)^{n^*-i} \left(iX - \beta^c n^* \delta_I\right) \\
&\quad + \sum_{i=\frac{\delta_I}{X}(1=\alpha)n^*+1}^{n^*} \binom{n^*}{i} P_B^i \left(1-P_B\right)^{n^*-i} \left(iX - \beta^c n^* \delta_I\right) \\
&< E\left(\pi_t | \left(\beta N, 0\right)\right),
\end{aligned}$$

where the last inequality follows since the second and third term equals the expected return on bad projects in n^* economies. Consequently, it is optimal for the international bank to finance only borrowers with good projects if $N \geq N^{**}$. ■

Lemma 12 *If $\Pr(P_I = P_G) > 0$, then there exists a number of markets, $N^* \geq N^{**}$, such that the foreign bank can always raise deposits when it is active in more than N^* markets.*

Proof. Define $\beta' \equiv \Pr(P_I = P_G)$. For $N \geq N^{**}$, the foreign bank finances borrowers with bad projects only if $\alpha = 0$. Let γ_t^k be the public signal observed by borrowers in economy k. Then, the probability that borrowers in economy k assign to the event $\alpha = 0$ is

$$\Pr\left(\alpha = 0 | \gamma_t^k < \gamma^*\right) = \Pr\left(P_I = P_B | \gamma_t^k\right) \left[\Pr\left(P_I = P_B\right)\right]^{N-1}.$$

Let the rate at which the foreign bank lends to borrowers be given by $\tilde{R}$. Conjecture that in equilibrium, $P_G \tilde{R} > R_f$. That indeed this is the case is verified in Proposition 9. Consider a state where $\gamma_t^k < \gamma^*$. Let n^* be as defined in the previous lemma, and let N' be the number of economies in which the foreign bank must be active to attract deposits from depositors in economy k. Then N' solves

$$\left[1 - \Pr\left(\alpha = 0|\gamma_t^k\right)\right] \times \left[\sum_{i=0}^{\frac{\delta_I}{\tilde{R}}\beta' N} \binom{\beta' N}{i} P_G^i \left(1 - P_G\right)^{\beta' N - i} \frac{i\tilde{R}}{\beta' N} + \sum_{i=\frac{\delta_I}{\tilde{R}}\beta' N+1}^{\beta' N} \binom{\beta' N}{i} P_G^i \left(1 - P_G\right)^{\beta' N - i} \delta_I\right] + \Pr\left(\alpha = 0|\gamma_t^k\right) \left[\sum_{i=0}^{\frac{\delta_I}{\tilde{R}}n^*} \binom{n^*}{i} P_B^i \left(1 - P_B^i\right)^{n^* - i} \frac{i\tilde{R}}{n^*} + \sum_{i=\frac{\delta_I}{\tilde{R}}n^*+1}^{n^*} \binom{n^*}{i} P_B^i \left(1 - P_B\right)^{n^* - i} \delta_I\right] \geq R_f,$$

where $\frac{i\tilde{R}}{\beta' N}$ and $\frac{i\tilde{R}}{n^*}$ are the depositors repayment when the return on the bank's portfolio is insufficient to repay the depositors in full. The maximal deposit rate that the bank can credibly promise its depositors is equal to the rate it charges its borrowers, so $\delta_I \leq \tilde{R}$. Let $\delta_I = \tilde{R}$. Thus, N' is the lowest value of N which ensures,

$$\left[1 - \Pr\left(\alpha = 0|\gamma_t^k\right)\right] \left[\sum_{i=0}^{\beta' N} \binom{\beta' N}{i} P_G^i \left(1 - P_G\right)^{\beta' N - i} \frac{i\tilde{R}}{\beta' N}\right] + \Pr\left(\alpha = 0|\gamma_t^k\right) \left[\sum_{i=0}^{n^*} \binom{n^*}{i} P_B^i \left(1 - P_B^i\right)^{n - i} \frac{i\tilde{R}}{n^*}\right] \geq R_f \quad (8)$$

$\lim\limits_{N\to\infty} \Pr\left(\alpha = 0|\gamma_t^k < \gamma^*\right) = 0$, so for $N \to \infty$, the left hand side of (8) goes to

$$\sum_{i=0}^{\beta' N} \binom{\beta' N}{i} P_G^i \left(1 - P_G\right)^{\beta' N - i} \frac{i\tilde{R}}{\beta' N}$$

$$\frac{\beta' N P_G \tilde{R}}{\beta' N} = P_G \tilde{R} > R_f.$$

$\lim\limits_{N\to 1} \Pr\left(\alpha = 0|\gamma_t^k < \gamma^*\right) = \Pr\left(P_t = (P_B, P_B)\,|\gamma_t^k < \gamma^*\right)$, so for $N \to 1$, the left hand side of (8) goes to

$$\Pr\left(P_t \neq (P_B, P_B)\,|\gamma_t^k\right) P_B \tilde{R} + \left[1 - \Pr\left(P_t = (P_B, P_B)\,|\gamma_t^k\right)\right] P_G \tilde{R} < R_f,$$

where the inequality follows since $\gamma_t^k < \gamma^*$. Thus, for any given realisation of the public signal, γ_t^k, there is a value of N, $N^* = \min\left(N', N^{**}\right)$, such that depositors in economy k are willing to finance the foreign bank. Note, that if $\gamma_t^k \geq \gamma^*$, the depositors will finance both the foreign and the local bank. ■ ■

Proof. Lemma 4. The incentive and compatibility constraints of the optimal collateral contract are given by,

$$P_B(X-R)-(1-P_B)C \leq 0,$$
$$P_G(X-R)-(1-P_G)C \geq 0,$$
$$(P_G R - R_f)+(1-P_G)\Delta C \geq 0,$$

where the third equality exploits that, when the bank's loans are subject to collateral, the bank's real cost of funds is R_f. Under the contract which maximizes the surplus of the borrowers with good projects, the first and third constraint are binding, so

$$\begin{aligned} R &= \frac{(1-P_B)R_f-(1-P_G)P_B\Delta X}{(1-P_B)P_G-(1-P_G)P_B\Delta}, \\ C &= \frac{(P_G-P_B)R_f+(1-P_G)P_G P_B(1-\Delta)X}{(1-P_B)P_G-(1-P_G)P_B\Delta}. \end{aligned}$$

Indeed, this contract fulfils the participation constraint of borrowers with good projects. If the first equality if fulfilled with equality,

$$C=\frac{P_B}{1-P_B}(X-R),$$

and therefore

$$\begin{aligned} & P_G(X-R)-(1-P_G)C \\ = \; & P_G(X-R)-\frac{1-P_G}{1-P_B}P_B(X-R) \\ = \; & \left(P_G-\frac{1-P_G}{1-P_B}P_B\right)(X-R)>0. \end{aligned}$$

For a borrower with a good project, the expected cost of a collateralised loan is

$$P_G R+(1-P_G)C=\frac{(P_G-P_B)R_f+(1-P_G)P_G P_B(1-\Delta)X}{(1-P_B)P_G-(1-P_G)P_B\Delta}.$$

Thus, if a bank offers the optimal collateral contract, and the competitor can observe the quality of the borrowers' projects, then the competitor can offer the contract $\left(\tilde{R},0\right)$, where

$$\begin{aligned} P_G\tilde{R} &= P_G R+(1-P_G)C \iff \\ \tilde{R} &= \frac{1}{P_G}\frac{(P_G-P_B)R_f+(1-P_G)P_G P_B(1-\Delta)X}{(1-P_B)P_G-(1-P_G)P_B\Delta}. \end{aligned}$$

Borrowers with good projects weakly prefer this contract to the collateral contract. Let $E\left(\pi^{c}\right)$ denote the expected profits to the relationship lender when the competitor offers a collateralized loan. Then,

$$\begin{aligned} E\left(\pi^{c}\right) &= P_{G}\tilde{R}-R_{f} \\ &= \frac{\left(1-P_{G}\right)P_{B}\left(1-\Delta\right)\left(P_{G}X-R_{f}\right)}{\left(1-P_{B}\right)P_{G}-\left(1-P_{G}\right)P_{B}\Delta}>0, \end{aligned}$$

where the last inequality follows since $\Delta<1$. If a bank is the only active lender in the market and it observes only the borrowers type and the public signal, then the contract which is accepted by good borrowers and rejected by bad borrowers and which maximizes the bank's profits fulfils the constraints,

$$\begin{aligned} P_{G}\left(X-R\right)-\left(1-P_{G}\right)C &= 0, \\ P_{B}\left(X-R\right)-\left(1-P_{B}\right)C &< 0, \\ C &= \rho, \end{aligned}$$

where $\rho\rightarrow 0$. The first and the third equation yields,

$$C=\rho \text{ and } R=X-\frac{1-P_{G}}{P_{G}}\rho.$$

Since $P_{B}<P_{G}$, so this contract ensures that borrowers with bad project reject the contract. For $\rho\rightarrow 0$, this contract allows the lender to extract the monopoly profits. This verifies the statements in the lemma. ■

Proof. Lemma 6. Condition (3) follows from $R_{1I}^{D}\geq R_{1D}^{D}$ and condition (4) follows from $R_{1D}^{I}\geq R_{1I}^{I}$. The proof follows directly from the discussion in the text combined with the observation that the local bank never poaches type I borrowers in period two. The latter observation is proved here. By Lemma 1, if the local bank could observe the type I borrowers project, it lends only to type I borrowers if $P_{I}=P_{G}$. Thus, the lower bound on the return from poaching type I borrowers is given by the return the local bank would obtain if it observed $P_{I}=P_{G}$. In equilibrium, $\delta_{D}>\frac{R_{f}}{P_{G}}$. In period two, the foreign bank is willing to lend to type I borrowers with good projects at $\frac{R_{f}}{P_{G}}$. Thus, to poach type I borrowers with good projects, the local bank must offer to lend below $\frac{R_{f}}{P_{G}}$.

Consequently, even if the local bank knew that type I borrowers held good projects and type I borrowers succeeded, the bank would make a loss from lending to type I borrowers. Thus, the upper bound on the local bank's returns from financing type I borrowers is negative, so the local bank cannot deviate from the equilibrium strategy. ■

Proof. Lemma 7. The foreign bank's expected profits from offering the optimal collateral contract in period one, and offering the optimal monopoly contract in period two is given by

$$\Pr\left(P_I = P_G \cap \gamma < \gamma^*\right)\left(P_G X - R_f\right).$$

The expected profits from offering the pooling contract to type I borrowers in period one and lending to type I borrowers with good projects in period two is,

$$\left(\bar{P}R^I_{1D} - R_f\right) + \left[\Pr\left(P_I = P_G \cap \gamma \geq \gamma^{**}\right)\pi^c + \Pr\left(P_I = P_G \cap \gamma < \gamma^{**}\right)\left(P_G X - R_f\right)\right].$$

The foreign bank offers the pooling contract if

$$\left(\bar{P}R^I_{1D} - R_f\right) + \left[\Pr\left(P_I = P_G \cap \gamma \geq \gamma^{**}\right)\pi^c + \Pr\left(P_I = P_G \cap \gamma < \gamma^{**}\right)\left(P_G X - R_f\right)\right] >$$
$$\Pr\left(P_I = P_G \cap \gamma < \gamma^*\right)\left(P_G X - R_f\right) \Leftrightarrow$$

$$\left(1 - \frac{\bar{P}}{P_G}\right)\left(\Pr\left(P_I = P_G \cap \gamma \geq \gamma^{**}\right)\pi^c - R_f\right) +$$
$$\left(P_G X - R_f\right)\left[\Pr\left(\gamma < \gamma^{**} \cap P_I = P_G\right) - \Pr\left(\gamma < \gamma^* \cap P_I = P_G\right)\right] \geq 0,$$

which yields the condition in the lemma. ■

Proof. Lemma 8. If $\Pr\left(P_D = P_B|\gamma^{**}\right) \leq \Pr\left(P_t = \left(P_B, P_B\right)|\gamma^*\right)$ then $\gamma^{**} > \gamma^*$. By (2) and (6),

$$\left[\Pr\left(P_D = P_B|\gamma^{**}\right)P_B + \left(1 - \Pr\left(P_D = P_B|\gamma^{**}\right)\right)P_G\right]\tilde{R}$$
$$= \left[\Pr\left(P_t = \left(P_B, P_B\right)|\gamma^*\right)P_B + \left(1 - \Pr\left(P = \left(P_B, P_B\right)|\gamma^*\right)\right)P_G\right]X \Leftrightarrow$$

$$\Pr\left(P_D = P_B|\gamma^{**}\right)P_B + \left(1 - \Pr\left(P_D = P_B|\gamma^{**}\right)\right)P_G >$$
$$\Pr\left(P_t = \left(P_B, P_B\right)|\gamma^*\right)P_B + \left(1 - \Pr\left(P_t = \left(P_B, P_B\right)|\gamma^*\right)\right)P_G \Leftrightarrow$$

$$\Pr\left(P_D = P_B|\gamma^{**}\right) < \Pr\left(P_t = \left(P_B, P_B\right)|\gamma^*\right).$$

Thus, $\gamma^{**} > \gamma^*$. ■

8 Literature

Allen, F., Carletti, E. and Marquez, R. (2008), "Credit Market Competition and Capital Regulation", *Robert H. Smith School Research Paper No. RHS 06-37*

Allen, F. and Gale, D. (2004), "Competition and Financial Stability", *Journal of Money Credit and Banking, Vol. 36, No. 3*

Bensako, B. and Thakor, A. (1987), "Competitive Equilibrium in the Credit Market under Asymmetric Information", *Journal of Economic Theory, Vol. 42, No. 1*

Berlin, M. and Mester, L. (1999), "Deposit and Relationship Lending", *Review of Financial Studies, Vol. 12, No. 3*

Berger, A., Klapper, L. and Udell, G. (2001), "The Ability of Banks to Lend to Informationally Opaque Small Businesses", *Journal of Banking and Finance, Vol. 25, No. 12*

Berger, A., Miller, N., Petersen, M., Rajan, R., and Stein, J. (2005), "Does Function Follow Organizational Forms? Evidence from the Lending Practices of Large and Small Banks, *Journal of Financial Economics, Vol. 76, No. 2*

Berger, A., Rosen, R., Udell, G. (2007), "Does Market Size Structure Affect Competition? The Case of Small Business Lending", *Journal of Banking of Finance 31, Vol. 31, No. 1*

Berger, A., and Udell, G. (2001) "Small Business Credit Availability and relationship Lending: The Importance of Bank Organisational Structure, *Economic Journal 112*

Berger, A. and Udell, G. (2006), "A more Complete Conceptual Framework for SME Finance", *Journal of Banking and Finance, Vol. 30, No. 11*

Bester, H. (1985), "Screening vs. Rationing in Credit Markets with Imperfect Information", *American Economic Review, Vol. 75, No. 4*

BIS (2004), "Foreign Direct Investment in the Financial Sector of Emerging Market Economies", *CGFS Papers No. 22, Bank for International Settlements*

BIS (2005), "Foreign Direct Investment in the Financial Sector - Experiences in Asia, Central and Eastern Europe and Latin America", *CGFS Papers No. 25, Bank for International Settlements*

Booth, A., and Thakor, A. (2000), "Can Relationship Banking Survive Competition?", *Journal* *of Finance Vol. 55, No. 2*

Booth, A., and Marinč, M. (2007), "Competition and Entry in Banking: Implications for Capital Regulation", *Universiteit von Amsterdam Working Paper*

Broecker, T. (1990), "Credit-Worthiness Tests and Interbank Competition", *Economectica Vol. 58, No. 4*

Cerasi, V. and Daltung, S. (2000), "The Optimal Size of a Bank: Costs and Benefits of Diversification", *European Economic Review, Vol. 44*

Claessens, S., Demirgüç-kunt, A. and Huizinga, H. (2001), "How does Foreign Entry Affect Domestic Banking Markets?", *Journal of Banking and Finance, Vol. 25, No. 5*

Clarke, G., Cull, R. and Soledad Martinez Peria, M. (2002), "Does Foreign Bank Penetration Reduce Access to Credit in Developing Countries? Evidence from Asking Borrowers", *World Bank Working Paper (February 2002)*

Clayes, S. and Hainz, C. (2006), "Acquisition versus Greenfield", *ECB Working Paper Series No.653*

Demirgüç-Kunt, A. and Huizinga, H. (1999), "Determinants of Commercial Bank Interest Margins and Profitability Some International Evidence", *World Bank Economic Review, Vol. 13, No. 2*

Demirgüç-kunt, A., Karacaovali, B., and Laeven, L. (2005), "Deposit Insurance around the World: A Comprehensive Database", *World Bank Policy Research Paper, No. 3628*

Dell'Ariccia, G., Marquez, R. (2004), "Information and Bank Credit Allocation", *Journal of Financial Economics, Vol. 72, No. 1*

Diamond, D. and Dybvig, P. (1983), Bank Runs, Deposit Insurance and Liquidity, *Journal of Political Economy, Vol. 91, No. 3*

Diamond, D. (1984), "Financial Intermediation and Delegated Monitoring", *Review of Economic Studies Vol. 51, No. 3*

Froot, K., Sharfstein, D. and Stein, J. (1993), "Risk Management: Coordinating Risk Management and Financial Policies", *Journal of Finance, Vol. 48, No. 5, 1993*

Gianetti, M. and Ongena, S. (2005): "Financial Integration and Entrepreneurial Activity: Evidence from Foreign Bank Entry in Foreign Markets", *Stockholm School of Economics Working Paper No. 05/02*

Galac, T. and Kraft, E. (2000), "What Has Been the Impact of Foreign Banks in Croatia?", *Croatian National Bank Survey No. S-4*

Goldberg, L. Dages, B. and Kinney, D. (2000), "Foreign and Domestic Banks Participation Emerging Markets: Lessons From Mexico and Argentina", *NBER Working Paper No. 7714*

Goldberg, L. (2001), "When is US Lending to Emerging Markets Volatile", *FRB of New York Staff Report No. 119*

Haas, R. de , and Lelyveld, I. van (2006), "Foreign Banks and Credit Stability in Central and Eastern Europe. A Panel Data Analysis", *Journal of Banking and Finance Vol. 30, No. 7*

Haas, R. de and Lelyveld, I. van (2006b), "Internal Capital Markets and Lending by Multinational Banks Subsidiaries", *De Nederlandsche Bank, Working Paper No. 101*

Haas, R. de and Naaborg, I (2005a), "Does Foreign Bank Entry Reduce Small Firms; Access to Credit? Evidence from European Transition Economies", *De Nederlandsche Bank, Working Paper No. 20*

Haas, R. de and Naaborg, I. (2005b), "Internal Capital Markets in Multinational Banks: Implications of European Transition Countries", *De Nederlandsche Bank, Working Paper No. 51*

Hauswald, R. and Marquez, R. (2006), "Competition and Strategic Information Acquisition in Credit Markets", *Review of Financial Studies, 2006, Vol. 79*

Kaminsky, G. (2005), "International Capital Flows: A Blessing or a Curse?", *UN/DESA Working Paper No. 10*

Konopielko, L. (1999), Foreign Banks' Entry into Central and East European Markets: Motives and Activities, *Post-Communist Economies, Vol. 11, No. 4, 1999*

Mian, A. (2006), "Distance Constraints: The Limits of Foreign Lending in Poor Economies", *Journal of Finance, Vol. 61, No. 3*

Mian, A. and Khwaja, A. (2006), "Tracing the Impact of Bank Liquidity Shocks: Evidence From an Emerging Market", *NBER Working Paper No. 12612*

Peek, J. and Rosengren E. (1997), "The International Transmission of Financial Shocks: The Case of Japan", *American Economic Review, Vol. 87*

Peek, J. and Rosengren, E. (2000), "Implications of the Globalization of the Banking Sector: The Latin American Experience", *New England Economic Review (September/October)*

Peria, M. Powell, A., Vladkova Hollar, I. (2002), "Banking on foreigners: The Behavior of International Bank Lending to Latin America", *World Bank Working Paper No. 2893*

Peria, M. and Mody, A. (2004), "How Foreign Participation and Market Concentration Impact Bank Spreads: Evidence from Latin America", *Journal of Money, Credit and Banking, Vol. 36, No. 3*

Stein, J., (2002), "Information Production and Capital Allocation: Decentralized vs. Hierarchical Firms", *Journal of Finance, Vol. 57, No. 5*

Sengupta, R. (2007), "Foreign Entry and Bank Competition", *Journal of Financial Economics, Vol. 84, No. 2 (2007)*

Williamson, O. (1988), "Corporate Finance and Corporate Governance", *Journal of Finance Vol. 43, No. 3*

Productivity, Preferences and UIP deviations in an Open Economy Business Cycle Model*

Arnab Bhattacharjee† Jagjit S. Chadha‡ Qi Sun§

June 2008

Abstract

We show that a flex-price two-sector open economy DSGE model *can* explain the poor degree of international risk sharing and exchange rate disconnect. We use a suite of model evaluation measures and examine the role of (i) traded and non-traded sectors; (ii) financial market incompleteness; (iii) preference shocks; (iv) deviations from UIP condition for the exchange rates; and (v) creditor status in net foreign assets. We find that there is a good case for both traded and non-traded productivity shocks as well as UIP deviations in explaining the puzzles.

JEL Classification: E32; F32; F41.

Keywords: Current account dynamics, real exchange rates, incomplete markets, financial frictions.

*Acknowledgements: We are grateful to seminar participants at the Centre for Dynamic Macroeconomic Analysis, the ESRC Evolving Macroeconomics Seminar, the Society of Computational Economics in Amsterdam 2004, the Money, Macro, Finance Conference 2004, 'Exchange Rates: Choice and Consequences', Cambridge 2007 and the 'Cycles, Contagion and Crises Workshop' June 2007 at the Financial Markets Group. We are grateful to Christoph Thoenissen for sharing the early results of his open economy work with us. Qi Sun thanks the Centre for Dynamic Macroeconomic Analysis for financial support and assistance.

†School of Economics and Finance, Castlecliffe, The Scores, St Andrews University, UK. E-mail: ab102@st-andrews.ac.uk.

‡Professor of Economics at the Department of Economics at Kent University at Canterbury. Professorial Fellow of the Centre for International Macroeconomics and Finance, Cambridge University. E-mail: jsc@kent.ac.uk.

§Centre for Dynamic Macroeconomic Analysis, School of Economics and Finance, Castlecliffe, The Scores, St Andrews University, UK. E-mail: qs@st-andrews.ac.uk.

1. Introduction[1]

It is well documented that international risk sharing and the real exchange rate seem to divert far from the levels that would be associated with their complete market allocations. Many authors, originating with Backus and Smith (1993) and Backus, Kehoe and Kydland (1995),[2] have pointed to a lack of aggregate risk sharing across open economies and as an analogue many have also commented on the disconnect between the relative price of goods and their relative consumption, see, for example, Obstfeld and Rogoff (2000) for a summary. We concentrate on a flexible price solution to the problem in the vein on Baxter and Crucini (1995) and Stockman and Tesar (1995) but also allow for financial market imperfections, following Devereux and Engel (2002). We find, within the context of a new methodology for model evaluation of calibrated models, that a two-sector open economy replete with financial market imperfections and driven by productivity, preference and exchange rates that are allowed to deviate stochastically from UIP may provide a reasonably satisfactory contribution to the solution of these puzzles.

To understand the puzzles, we use a two-sector version of Chari, Kehoe and McGrattan (2002), developed by Benigno and Thoenissen (2004), in which there are infinitely-lived representative optimizing households, a two-sector production sector for traded and non-traded goods, where the law of one price holds but where there are also incomplete financial markets. As is well known, under a complete markets environment, cross-country holdings of assets should be sufficient to ensure that consumption rather than income is highly correlated in open economies and that relative consumption responds to changes in relative prices.[3] Because considerable evidence has suggested that international portfolios

[1]This paper has been presented at St Andrews University, Aberdeen University, Brunel University, London Metropolitan University, the Cass Business School, University College, Oxford, Kent University, LSE, Reading University and at the Bank of Iceland and the Norges Bank. We thank Farooq Akram, Michael Bordo, Ehsan Choudhri, John Driffill, Charles Goodhart, Chris Meissner, Marcus Miller, Charles Nolan, Joe Pearlman, Lucio Sarno, Katsuyuki Shibayama, Alan Sutherland, Ashley Taylor, Mark Taylor and Simon Wren-Lewis for helpful comments.

[2]Simply put the Backus-Kehoe-Kydland puzzle is that it is income rather than consumption that is more closely correlated across open economies, which suggests that payoffs from idiosyncratic foreign (domestic) income shocks are not being used to smooth domestic (foreign) consumption. The Backus-Smith puzzle is the analogous puzzle that relative consumption across open economies does not arbitrage relative price (real exchange rate) differences.

[3]Baxter and Jermann (1997) conclude, with a wealth holding model with a production sector, that domestic individuals should hold only foreign shares against loss caused for labour income

are home-biased (Tesar and Werner, 1995) and imply that an important channel for risk sharing may be impeded, to some extent, a popular treatment is to introduce incomplete markets by assuming that portfolio diversification relies only on non-state contingent bonds, as in Kehoe and Perri (2002), and accordingly we adopt this feature.[4]

Full price flexibility is maintained in the model but real rigidities are present in the form of a home bias in both consumption and the use of both traded and non-traded goods in output. The model we adopt also allows for costly capital accumulation, an interest rate spread and the possibility of a country being a net creditor (or debtor).[5] The model is driven by three types of shocks: to both traded and non-traded sector productivity; to preferences in the allocation of time between work and leisure of the representative household, and by deviations of the exchange rate from the path expected by relative interest rates (see, Frankel, 1996, and Sarno and Taylor, 2002).

A further innovation of this paper is the development of summary statistics on the distance of each model simulation to the data in the sense of Geweke's (1999) 'weak' interface with the data. We define a model as a structural set of equations, which are parameterized, and simulated with forcing variables defined over a given variance-covariance matrix (VCM) of shocks. The model then produces an artificial economy which can be thought of as lying some distance from our systematic observations on real-world economies (Watson, 1993). In this sense, the open-economy puzzles drive a large wedge between theory and observation and so we construct a number of empirical measures of this wedge across models and choice of forcing variables to understand which models provide a more satisfactory resolution of the puzzles (see, for example, Ledoit and Wolf, 2002, for related work).

Our results suggest that some form of financial market incompleteness will probably be required to solve the open-economy puzzles (as suggested by Engel, 2000). A key result is that price stickiness may not necessarily be required to resolve the puzzles. It turns out that reasonable answers can be found with reference to traded and non-traded forcing processes and by allowing the exchange

by a domestic negative shock.

[4]Recently authors such as Sorensen *et al* (2007) have documented a reduction in home bias but continue to draw a clear link between home bias and risk sharing. Our set-up is sufficiently flexible to allow us to alter the cost of borrowing from abroad.

[5]The importance of these creditor or debtor positions have been explored comprehensively by Lane and Milesi-Ferretti (2002).

rate to deviate from the UIP condition. In the former case, with a dominant role for traded over non-traded productivity shocks, in an incomplete financial market, domestic households raise consumption for traded and non-traded goods compared to overseas but the real exchange rate depreciates if the terms of trade effect outweighs the Harrod-Balassa-Samuelson effect (Corsetti *et al*, 2004). In the case of preference (for work over leisure) shocks, the labor supply curve shifts out and hence demand for goods increases (Hall, 1997) but with an elastic investment supply schedule, and hence output, there is little response in the real exchange rate. And deviations from the uncovered interest rate parity equation for the exchange rate can operate to drive the exchange rate to appreciate even if domestic interest rates fall. Consumption increases in response to the fall in real rates and investment also increases, with wage growth attenuated by the exchange rate appreciation and this results in a reduction in net foreign assets (a current account deficit). Finally, it can also be shown that a combination of these shocks seems to explain the puzzles best.

1.1. Some simple observations

We examine open economy data from 24 OECD and emerging country economies. Figure 1 gives the descriptive statistics of HP filtered cyclical data and illustrates some clues that the behavior of the current account over the cycle is likely to help explain the puzzles. We note that (i) the real exchange rate is considerably more volatile than relative consumption; (ii) that relative output still seems more correlated than relative consumption; (iii) that current and trade account dynamics follow each other closely and (iv) that the current account is (mostly) countercyclical.

Figure 1 is set over four panels. The top left hand panel of Figure 1 shows the extent to which the real exchange rate seems noisy and significantly more volatile than its fundamentals would imply. The range for observed volatility of the real exchange rate is between 1-9, with an average, over this dataset of nearly 4. Researchers have explained this high volatility from many dimensions in the literature.[6] And certainly, we find that compared to relative consumption, which ranges from 0.5 to just under 3, the real exchange rate does look 'disconnected'. The top right hand side panel of Figure 1 scatters the correlation of national consumption with US consumption of the economies against the correlation of

[6] These explanations include price stickiness and the famous case of exchange rate overshooting (Dornbusch, 1976).

output with US output and suggests in general that output is more closely related across countries than consumption, which implies somewhat less than perfect risk sharing.

The left hand lower panel of Figure 1 shows the close correspondence between the business cycle dynamics of the current account and the trade balance over the business cycle across these economies - suggesting a strong role for intertemporal trade over the business cycle with some deviation from complete markets as the balance on the trade account is not offset by returns from assets held overseas.[7]

Finally, the lower right hand side panel of Figure 1 suggests that the current account tends to be countercyclical (with a deficit under an economic expansion). But that the real exchange rate looks as likely to appreciate or depreciate over the same economic cycle. Put alternatively, there is a higher demand for foreign assets during an expansion (with current account output correlations negative) but that the real exchange rate plays a limited role in choking off that higher demand.

A second modelling question concerns whether price stickiness is required for the resolution of the puzzles. Figure 2 shows the forecast error correlation of up to 25 quarters of US and UK current account and real exchange rate and relative consumption and the real exchange rate (den Haan, 2000). The panels show that over long run, these quantities are countercyclical but over the short term, all three measures somewhat less so. As price stickiness can be expected to play a less important role in long run dynamics, than in short run, there is some initial motivation for excluding this feature from our model.

The rest of the paper is organized as follows. Section 2 describes the model, section 3 outlines the solution technique and model calibration, section 4 offers the model results, section 5 compares the model to the data VCM and section 6 concludes. Appendices A and B offer more detail on model, shock selection and the evaluation methodology.

2. The Model

This section describes the baseline model. Essentially, we take the flexible price two-country, two sector model derived by Benigno and Thoenissen (2004) and

[7] The finding that the current account is likely to play an important role in the resolution of puzzles has two implications for our work, we will want to adopt a model where current account dynamics play an important role and assess the fit of any models we develop with, inter alia, their match to current account data.

emphasize the specification of driving forces as in Chadha, Janssen and Nolan (2001). The model is driven variously by forcing variables in domestic and overseas traded and non-traded productivity shocks, domestic and overseas preference shocks and by deviations from the UIP condition for the exchange rate.

2.1. Consumer behavior

We adopt a two-country model. Consumers are infinitely lived. The world economy is populated by a continuum of agents on the interval $[0,1]$, with the segment $[0,n)$ belonging to the country H (Home) and the population on segment $[n,1]$ belonging to the F (Foreign) country. Preferences for the Home consumer (with an identical set-up for the foreign consumer) are described by the utility function:

$$U_t = E_t \sum_{s=t}^{\infty} \beta^{s-t} \left[U(C_s^j, \xi_{C,s}) V(l_s^j) \right], \tag{2.1}$$

where E_t denotes the expectation conditional on the information set at date t, and β is the intertemporal discount factor, with $0 < \beta < 1$. The Home consumer obtains utility from consumption, C^j, and receives disutility from supplying labor, l^j. $\xi_{C,s}$ is a stochastic disturbance affecting the utility the agent receives from a unit of consumption.

The asset market structure in the model is standard and is described in detail in Benigno (2001) and Benigno and Thoenissen (2004). Home individuals are able to trade two nominal bonds denominated in the domestic and foreign currency. The bonds are issued by residents in both countries in order to finance their consumption expenditure. Foreign residents, on the other hand, can allocate their wealth only in bonds denominated in the foreign currency. Home households face a cost when they take a position in the foreign bond market. As in Benigno (2001), this transaction cost depends on the net foreign asset position of the home economy.[8]

The Home consumer maximizes utility subject to the following budget constraint:

$$P_t C_t^j + \frac{B_{H,t}^j}{(1+i_t)} + \frac{S_t B_{F,t}^j}{(1+i_t^*)\Theta\left(\frac{S_t B_{F,t}}{P_t}\right)} = B_{H,t-1}^j + S_t B_{F,t-1}^j + P_t w_t l_t^j + \Pi_t^j \tag{2.2}$$

[8] Alternative ways of closing open economy models are discussed in Schmitt-Grohe and Uribe (2003).

where P_t is the price index corresponding the basket of final goods C, w is the real wage earned by agent in return for supplying labor and Π are dividends received by the agent from holding an equal share of the economy's intermediate goods producing firms.

Home agents can hold two types of nominal, non-state contingent bonds. B_H^j denotes agent j's holdings of Home-currency denominated bonds. The one-period return from these bonds is denoted by $(1+i_t)$. S denotes the nominal exchange rate, defined as Home currency price of a unit of foreign currency. B_F^j denotes agent j's holdings of Foreign-currency denominated bonds. The one-period return from foreign-currency denominated bonds is $(1+i_t^*)\Theta\left(\frac{S_t B_{F,t}}{P_t}\right)$, where $(1+i_t^*)$ is the gross rate of return and $\Theta\left(\frac{S_t B_{F,t}}{P_t}\right)$ is a proportional cost associated with foreign currency-denominated bond holding that depends on the economy-wide holdings of foreign-currency denominated bonds.[9]

The first order condition of the representative consumer can be summarized as follows:

$$U_{c,t} = (1+i_t)\beta E_t\left[U_{c,t+1}\frac{P_t}{P_{t+1}}\right] \tag{2.3}$$

$$U_{c,t+1} = (1+i_t^*)\Theta\left(\frac{S_t B_{F,t}}{P_t}\right)\beta E_t\left[U_{c,t+1}\frac{S_{t+1}P_t}{S_t P_{t+1}}\right]. \tag{2.4}$$

$$U_{c,s}w_t = V_l(l_s) \tag{2.5}$$

where $U_{c,t} \equiv U_c(C_t, \xi_{C,t}, 1-l_t)$ and where there is an analogous intertemporal condition to (2.3) for the Foreign consumer. As in Benigno (2001), we assume that all individuals belonging to the same country have the same level of initial wealth. This assumption, along with the fact that all individuals face the same labor demand and own an equal share of all firms, implies that within the same country all individuals face the same budget constraint and so they will choose identical paths for consumption. As a result, we are able to drop the j superscript and focus on a representative individual for each country.

2.2. The supply side

In this economy there are three layers of production in this economy. Final goods are produced by a competitive final goods producing sector using Home traded and

[9]The factor of proportionality $\Theta\left(\frac{S_t B_{F,t}}{P_t}\right)$ is equal to unity only when economy-wide bond holdings are at their initial steady state level, thus ensuring that in the long-run the economy returns to its initial steady state level of bond holdings.

non-traded intermediate goods as well as foreign-produced traded intermediate-goods. Final goods are non-traded and are either consumed or used as investment goods to augment the domestic capital stock. Intermediate goods producers combine labor and capital according to a constant returns to scale production technology to produce intermediate goods. Each country produces two types of intermediate goods, a differentiated traded good and a non-traded good.

2.2.1. Final good producers

Let Y be the output of final good produced in the home country. Final goods producers combine domestic and foreign-produced intermediate goods which they must obtain from the distributor to produced Y in a two-step process. The final good Y is made up of traded, y_T, and non-traded inputs, y_{NT}, combined in the following manner:

$$Y = \left[\omega^{\frac{1}{\kappa}} y_T^{\frac{\kappa-1}{\kappa}} + (1-\omega)^{\frac{1}{\kappa}} y_N^{\frac{\kappa-1}{\kappa}}\right]^{\frac{\kappa}{\kappa-1}}, \tag{2.6}$$

where ω is the share of traded goods in the final good, and κ is the intratemporal elasticity of substitution between traded and non-traded intermediate goods. The traded component, $y_{T,}$ is, in turn, produced using home, y_H, and foreign-produced traded goods, y_F, in the following manner:

$$y_T = \left[v^{\frac{1}{\theta}} y_H^{\frac{\theta-1}{\theta}} + (1-v)^{\frac{1}{\theta}} y_F^{\frac{\theta-1}{\theta}}\right]^{\frac{\theta}{\theta-1}}, \tag{2.7}$$

where v is the domestic share of home produced traded intermediate goods in total traded intermediate goods and θ is the elasticity of substitution between home and foreign-produced traded goods. Final goods producers are competitive and maximize profits, where P is the aggregate or sectoral price index and Y aggregate output.

$$\max_{y_N, y_H, y_F} PY - P_T y_T - P_N y_N, \tag{2.8}$$

subject to (2.7), where traded goods' output is maximized subject to the value of home and foreign traded goods.

This maximization yields the following input demand functions for the home and foreign (not shown but identical) firm:

$$
\begin{aligned}
y_N &= (1-\omega)\left(\frac{P_N}{P}\right)^{-\kappa} Y \qquad (2.9)\\
y_H &= \omega v\left(\frac{P_H}{P_T}\right)^{-\theta}\left(\frac{P_T}{P}\right)^{-\kappa} Y\\
y_F &= \omega(1-v)\left(\frac{P_F}{P_T}\right)^{-\theta}\left(\frac{P_T}{P}\right)^{-\kappa} Y.
\end{aligned}
$$

The price index that corresponds to the above maximization problem is:

$$
\begin{aligned}
P_T^{1-\theta} &= [vP_H^{1-\theta} + (1-v)P_F^{1-\theta}] \qquad (2.10)\\
P^{1-\kappa} &= [\omega P_T^{1-\kappa} + (1-\omega)\, P_N^{1-\kappa}],
\end{aligned}
$$

And the goods produced in the final goods sector are only used domestically, either for consumption or investment, x_t, for home and overseas:

$$
Y_t = C_t + x_t. \qquad (2.11)
$$

2.2.2. Traded-intermediate goods sector

Firms in the traded intermediate goods sector produce goods using capital and labor services. The typical firm maximizes the following profit function:

$$
\max P_{H_t} y_{H_t} + S_t P^*_{H_t} y^*_H - P_t w_t l_{H,t} - P_t x_{H,t}, \qquad (2.12)
$$

or because the law of one price holds at the wholesale level,

$$
\max_{H_t} P_{H_t}\left(y_{H_t} + y^*_H\right) - P_t w_t l_{H,t} - P_t x_{H,t}.
$$

This maximization is subject to:

$$
\begin{aligned}
y_{H_t} + y^*_{H_t} &= F(k_{H,t-1}, l_{H,t}) = \left(A_t l_{H,t}\right)^{\alpha} k_{H,t-1}^{1-\alpha} \qquad (2.13)\\
k_{H,t} &= (1-\delta)k_{H,t-1} + x_{H,t} - \phi\left(\frac{x_{Ht}}{k_{Ht-1}}\right) k_{Ht-1}.
\end{aligned}
$$

The stochastic maximization problem of the domestic intermediate goods firm is given by:

$$
L = \mathrm{E}_t \sum_{t=0}^{\infty} \beta^t \frac{U_{c,t}}{P_t} \left\{ \begin{array}{c} \left[P_{H,t}\left(A_t l_t\right)^{\alpha} (k_{H,t-1})^{1-\alpha} - P_t w_t l_{H,t} - P_t x_{H,t}\right] \\ +\lambda_t \left[\begin{array}{c} (1-\delta)k_{H,t-1} + x_{H,t} \\ -\phi\left(\frac{x_{H,t}}{k_{Ht-1}}\right) k_{H,t-1} - k_{H,t} \end{array} \right] \end{array} \right\}. \qquad (2.14)
$$

The first order conditions with respect to the labor input, investment and capital are given by:

$$P_t w_t = \alpha P_{H,t}(A_t)^{\alpha}(\frac{k_{H,t-1}}{l_{H,t}})^{1-\alpha}, \tag{2.15}$$

$$P_t = \lambda_t - \phi'\left(\frac{x_{H,t}}{k_{H,t-1}}\right)\lambda_t, \tag{2.16}$$

$$\lambda_t = \mathrm{E}_t\beta\frac{U_{c,t+1}}{U_{c,t}}\frac{P_t}{P_{t+1}}\left\{\begin{array}{c} P_{H_{t+1}}(1-\alpha)\left(\frac{A_{t+1}l_{H,t+1}}{k_{Ht}}\right)^{\alpha}+ \\ \lambda_{t+1}\left[(1-\delta)-\phi\left(\frac{x_{Ht+1}}{k_{H,t}}\right)+\phi'\left(\frac{x_{H,t+1}}{k_{H,t}}\right)\frac{x_{H,t+1}}{k_{H,t}}\right]\end{array}\right\}. \tag{2.17}$$

And using the expression for $P_{H,t}$ from the wage equation yields:

$$\lambda_t = \mathrm{E}_t\beta\frac{U_{c,t+1}}{U_{c,t}}\frac{P_t}{P_{t+1}}\left\{\begin{array}{c} \frac{(1-\alpha)}{\alpha}\left(\frac{l_{t+1}}{k_t}\right)P_{t+1}w_{t+1}+ \\ \lambda_{t+1}\left[(1-\delta)-\phi\left(\frac{x_{H,t+1}}{k_{H,t}}\right)+\phi'\left(\frac{x_{H,t+1}}{k_{H,t}}\right)\frac{x_{H,t+1}}{k_{H,t}}\right]\end{array}\right\}.$$

Next, substitute in the expression for λ

$$\begin{aligned} U_{c,t} &= \left[1-\phi'\left(\frac{x_t}{k_{t-1}}\right)\right]\mathrm{E}_t\beta U_{c,t+1}w_{t+1}\frac{f_{k_{t+1}}}{f_{l_{t+1}}}+ \\ &\quad \mathrm{E}_t\beta\frac{1-\phi'\left(\frac{x_t}{k_{t-1}}\right)}{1-\phi'\left(\frac{x_{t+1}}{k_t}\right)}U_{c,t+1}\left[(1-\delta)-\phi\left(\frac{x_{H,t+1}}{k_{H,t}}\right)+\phi'\left(\frac{x_{H,t+1}}{k_{H,t}}\right)\frac{x_{H,t+1}}{k_{H,t}}\right], \end{aligned} \tag{2.18}$$

where f_{k_t} is the marginal product of capital and $f_{l_{t+1}}$ the marginal product of labor and w_{t+1} is the real wage, $U_{c,t} \equiv U_c(C_t, \xi_{C,t}, 1-l_t)$.

2.2.3. Non-traded-intermediate goods sector

The non-traded intermediate goods producer has the similar maximization problem:

$$\max P_{N_t}y_{N_t} - P_t w_t l_{N,t} - P_t x_{N,t}, \tag{2.19}$$

which is subject to

$$\begin{aligned} y_{N_t} &= F(k_{t-1,}l_{N,t}) \\ k_{N,t} &= (1-\delta)k_{N,t-1} + x_t - \phi\left(\frac{x_{N,t}}{k_{N,t-1}}\right)k_{N,t-1}, \end{aligned} \tag{2.20}$$

and where $\psi y_{H,t} + \psi y_{F,t}$ are demands for non-traded goods coming from the distribution sector. If we now set up the stochastic maximization problem of the domestic intermediate goods firm:

$$L = \mathrm{E}_t \sum_{t=0}^{\infty} \beta^t \frac{U_{c,t}}{P_t} \left\{ \begin{array}{c} \left[\begin{array}{c} P_{N,t} \left(A_{N,t} l_{N,t}\right)^{\alpha} \left(k_{N,t-1}\right)^{1-\alpha} + P_{N,t} \left(\psi y_{H,t} + \psi y_{F,t}\right) \\ -P_t w_t l_{Nt} - P_t x_{N,t} \end{array} \right] \\ +\lambda_t \left[\begin{array}{c} (1-\delta) k_{N,t-1} + x_{N,t} \\ -\phi\left(\frac{x_{N,t}}{k_{N,t-1}}\right) k_{N,t-1} - k_{N,t} \end{array} \right] \end{array} \right\}. \tag{2.21}$$

The first order condition with respect to labor input is then given by:

$$P_t w_t = \alpha P_{N,t} (A_{N,t})^{\alpha} (\frac{k_{N,t-1}}{l_{N,t}})^{1-\alpha}.$$

The first order condition with respect to investment is:

$$P_t = \lambda_t - \phi' \left(\frac{x_{N,t}}{k_{N,t-1}} \right) \lambda_t.$$

The first order condition with respect to capital is:

$$\lambda_t = \mathrm{E}_t \beta \frac{U_{c,t+1}}{U_{c,t}} \frac{P_t}{P_{t+1}} \left\{ \begin{array}{c} P_{N_{t+1}} (1-\alpha) \left(\frac{A_{t+1} l_{N,t+1}}{k_{N,t}} \right)^{\alpha} + \\ \lambda_{t+1} \left[(1-\delta) - \phi\left(\frac{x_{N,t+1}}{k_{N,t}}\right) + \phi'\left(\frac{x_{N,t+1}}{k_{N,t}}\right) \frac{x_{N,t+1}}{k_{N,t}} \right] \end{array} \right\}, \tag{2.22}$$

and using the expression for P_N from the wage equation yields:

$$\lambda_t = \mathrm{E}_t \beta \frac{U_{c,t+1}}{U_{c,t}} \frac{P_t}{P_{t+1}} \left\{ \begin{array}{c} \frac{(1-\alpha)}{\alpha} \left(\frac{l_{N_{t+1}}}{k_{N_t}} \right) P_{t+1} w_{t+1} + \\ \lambda_{t+1} \left[(1-\delta) - \phi\left(\frac{x_{N,t+1}}{k_{N,t}}\right) + \phi'\left(\frac{x_{N,t+1}}{k_{N,t}}\right) \frac{x_{N,t+1}}{k_{N,t}} \right] \end{array} \right\}.$$

We next substitute in the expression for λ

$$\begin{aligned} U_{c,t} \quad = \quad & \left[1 - \phi'\left(\frac{x_{N,t}}{k_{Nt-1}} \right) \right] \mathrm{E}\beta U_{c,t+1} w_{t+1} \frac{f_{k_{t+1}}}{f_{l_{t+1}}} + \\ & \mathrm{E}_t \beta \frac{1 - \phi'\left(\frac{x_{N,t}}{k_{N,t-1}}\right)}{1 - \phi'\left(\frac{x_{N,t+1}}{k_{N,t}}\right)} U_{c,t+1} \left[(1-\delta) - \phi\left(\frac{x_{N.t+1}}{k_{N,t}}\right) + \phi'\left(\frac{x_{N,t+1}}{k_{N,t}}\right) \frac{x_{N,t+1}}{k_{N,t}} \right], \end{aligned} \tag{2.23}$$

where f_{k_t} is the marginal product of capital and $f_{l_{t+1}}$ the marginal product of labor and w_{t+1} is the real wage, $U_{c,t} \equiv U_c(C_t, \xi_{C,t}, 1 - l_t)$.

2.3. The real exchange rate

In this model, the real exchange rate is defined as:

$$RS_t = \frac{S_t P_t^*}{P_t} \tag{2.24}$$

and can deviate from purchasing power parity (PPP) as a result of three channels. As in Benigno and Thoenissen (2004), allowing for the possibility of home bias in consumption ($v > v^*$), via the terms of trade channel (because of home bias) and via the internal real exchange rate channel (because of non-traded goods), (2.24) can be expanded to give:

$$\frac{S_t P_t^*}{P_t} = \frac{S_t P_{H,t}^*}{P_{H,t}} \frac{P_{H,t}}{P_{T,t}} \frac{P_{T,t}^*}{P_{H,t}^*} \frac{P_{T,t}}{P_t} \frac{P_t^*}{P_{T,t}^*},$$

which when linearized around the steady state, where $\frac{SP^*}{P}$ equals unity, can be shown to be equal to:

$$\widehat{RS}_t = (v - v^*)\hat{T}_t + (\omega - 1)\, \hat{R}_t + (1 - \omega^*)\, \hat{R}_t^*. \tag{2.25}$$

The deviation of the real exchange rate around its steady state depends on deviations of the home and foreign retail to wholesale price ratios, the terms of trade, T, defined as $\frac{P_F}{P_H}$, and the relative price of non-traded to traded goods, R.

2.4. The current account

The current account is defined as changes in foreign asset holding, within the incomplete financial market. Home and foreign agents trade intermediate goods and the trade balance is used to buy foreign bonds and so the flow budget constraint shows the current account dynamics below. The left hand side is the changes in foreign asset holding. The right hand side shows the total production (first two terms) minus consumption and investment, yielding adjustment of bond wealth:

$$\frac{S_t B_t^F}{P_t (1 + i_t^*)} \frac{1}{\Phi\left(\frac{S_t B_t^F}{P_t}\right)} - \frac{S_t B_{t-1}^F}{P_t} = \frac{P_{H_t}}{P_t} (y_{H_t} + y_H^*) + \frac{P_{N_t}}{P_t} y_{N_t} - C_t - x_t. \tag{2.26}$$

2.5. Forcing variables

We adopt the specification of Stockman and Tesar (1995) and Chadha, Janssen and Nolan (2001) by investigating the role of both productivity and preference shocks for an open economy. We use both traded sector and non-traded sector productivity, which drive the input and hence product price, shocks to the allocation of time spent in work over leisure, which affects labor supply, stochastic deviations in the UIP condition, which directly affects the terms of trade. Each shock originates from a different sector but allows us to attribute exchange rate volatility to more than one exogenous factor. In total, we enable seven shocks (two sectoral and a preference shock in each of two countries, plus UIP deviations) and try to locate the importance in explaining open economy business cycles. The construction of each shock process is explained in Appendix A.

3. Solution and Model Calibration

3.1. Solution method

Before solving the model, it is log-linearized around the steady state to obtain a set of equations describing the equilibrium fluctuations of the model. The log-linearization yields a system of linear difference equations which we list in an appendix and can be expressed as a singular dynamic system of the following form:

$$\mathbf{A}\mathrm{E}_t\mathbf{y}(t+1 \mid t) = \mathbf{B}\mathbf{y}(t) + \mathbf{C}\mathbf{x}(t)$$

where $\mathbf{y}(t)$ is ordered so that the non-predetermined variables appear first and the predetermined variables appear last, and $\mathbf{x}(t)$ is a martingale difference sequence. There are up to seven shocks in $\mathbf{C}$. The variance-covariance as well as the autocorrelation matrices associated with these shocks are described in Table 1. Given an initial parametrization of the model, which we describe in the next section, we solve this system using the King and Watson (1998) solution algorithm.

3.2. Data and estimation

Table 1 summarizes the calibration parameters for the baseline simulation of the model. We collect both quarterly and annual data and calibrate the model for UK-US. Values of parameters are either estimated from US or UK data or taken from extant literature. An annual risk free rate of 4% and depreciation of 10% is assumed. Labor share is 0.6 and 0.577 for UK and US annual data. We take the consumption and leisure curvature of 2 (Corsetti *et al*, 2004) and 4 (Chadha *et al*, 2001). The elasticity of substitution between home and foreign goods in UK is 1.5 as in Chari *et al* (2002). For the trade-off between traded and non-traded goods we adopt the elasticity suggested by Corsetti *et al* (2004) of 0.74. UK and US trade data reveals the shares of UK produced goods in UK and US production to be 0.73 and 0.0157. Traded goods weights in all household consumption are estimated to be 0.3 and 0.24, smaller than that of Corsetti *et al* (2004), 0.45 to 0.5. Cost of financial intermediation is 4bp as in Benigno (2001). The cost of investment, $b = 2$, is chosen to match the relative volatility of investment.[10] Steady state of net foreign asset is set to be 0 or 0.2 which means, respectively, that the UK has a balanced current account or is a creditor.

We have at most seven exogenous shocks in our experiments. The vector of shocks Π_t are assumed to follow a VAR(1) process:

$$\Pi_{t+1} = A\Pi_{t+1} + U_t,$$

$$U_t \sim N(0, \Sigma).$$

4. Model Results

We now turn to the evaluation of the structural linear model by its simulation and comparison to our observations on the economy. As well as standard matching of moments, we develop a new approach for model evaluation and model selection.

4.1. Methodology

Conventional tools such as the impulse response function and variance decomposition help us understand the dynamics of an artificial economy. The standard practice is also to assess models against some selected second moments of the data. But in this paper we introduce criteria that takes into account all the second moments and evaluate model performance based on formal statistical measures. We define a better model, as one that can render a better match between VCM of the data and the VCM simulated by the model. In order to pin down some parameter value or decide on certain features of a model, we work on a class of candidate models (or calibrations). By examining the corresponding

[10] However we run experiments with also b=5, which are available on request and covered in the robustness exercise of Figure 7.

match for candidate models, we call any improvement towards the criterion a gain in marginal information. We can also evaluate the gain on a particular parameter, by which we can signal the importance of any one feature of the model. Strictly speaking, we cannot guarantee the marginal information gain is reliable, or nearer to the 'true' model, unless we are quite certain about the rest of the model. The proposition of a marginal information gain we make is therefore a 'weak-form' of model selection (see Geweke, 1999).

The criteria we use involve the statistical divergence of the two VCMs. We develop formal and also intuitive distance measures elsewhere but some details are available in Appendix B. A higher value of distance denotes a model that is further from our measure on 'true' data process.[11] The data required to evaluate the open economy model is of high dimension and a relatively short sample, which tends to make the model evaluation and selection very difficult. We calculate for each candidate model a distance and compare across each measure. We are cautious in making a proposition of model selection, especially for a particular parameter constellation, but feel able to make some statements on the validity of the joint choices on model and shock processes.

4.2. Impulse responses

The impulse response functions are based on the seven-shock model. In this calibration, the foreign country has the same properties as home, such as shares of traded and home goods on market. We change v=0.85 and v*=0.15, the home produced share of tradeables in intermediate goods production, in order to highlight the effect of foreign sector.

4.2.1. Traded productivity shocks

Figure 3 plots the response of quantities and relative prices to a traded productivity shock in the home country. The response of real exchange rate depends on two effects: the terms of trade and the Harrod-Balassa-Samuelson (HBS) effect. The former requires an adjustment in relative traded prices, which requires a depreciation in the real exchange rate in the long run. But the latter effect drives up wages in both the traded and non-traded sector but with no

[11] In developing this approach, we use Monte Carlo simulations on some artificial models. We find: (1) this approach works very well for models close to 'true model', as long as the multivariate normality is tenable; (2) our approach helps overcome small-sample bias, and (3) experiments on a sub-block of the full VCM may be inconclusive.

productivity improvement in the non-traded sector, non-traded prices will rise and hence so will the real exchange rate. This effect is especially strong, see section 2.3, when there is a home bias in consumption, which acts to accentuate the real exchange rate change. Finally, the lack of complete risk sharing means that consumption is more elastic to a productivity shock than under a complete markets allocation. The combination of forward-looking domestic consumption responding to higher productivity (income) but an attenuated overall investment response - where traded sector investment rises but non-traded sector investment falls - leads to the accumulation of foreign debt to finance current demand.

4.2.2. Non-traded productivity shocks

Following a non-traded productivity shock (Figure 4), investment and labor increase. Home households enjoy somewhat higher consumption in this case, more so than in the case of traded sector productivity shock. In this case, the terms of trade effect and HBS effect are the same, causing the real exchange rate to depreciate. Although the response of relative consumption is positive, it is not large enough to bring about a current account deficit, because there is a larger response from the labor input, and hence there is net lending overseas. In general the impulse responses suggest that strong traded-sector productivity shocks can lead to the matching of some elements of the open economy. A lack of complete risk sharing raises consumption at home compared to abroad and a strong preference for home goods consumption also amplifies the extent to which output increases.

4.2.3. Preference shocks

In principle, preference shocks might be thought to contribute a solution to the Backus-Smith puzzle simply as marginal utility is now, *inter alia*, a function of the preference shocks rather than just consumption growth:

$$RS = \frac{U_C^*}{U_C},$$

where we note that in the real exchange rate can be thought of as related to the ratio of marginal utilities in consumption (in a complete markets set-up). But by themselves may not provide a resolution as they seems to imply relatively acyclical current account dynamics and a reduction of real exchange rates along with higher domestic supply (see Chadha *et al*, 2001). This is because preference shocks alter the equilibrium point in the household trade-off between leisure and consumption.

Following Hall (1997) such shocks simply suggest that the household decides to allocate more (or less) time to work, which finances consumption, rather than leisure. As one would expect preference shocks help increase the volatility of the labor input by introducing exogenous shifts in work and may act to solve the puzzle of the Backus-Smith correlation (Figure 5). A home preference shock drives up labor input and consumption and reduces relative prices, if the supply response is elastic. So unless home agents become elastic in the substitution of leisure across periods, increased consumption is also met by an increase in investment and the current account remains acyclical.

4.2.4. Stochastic deviations from UIP

Following the suggestion of Devereux and Engel (2002), we explore the implication of stochastic deviations from the uncovered interest rate parity (UIP) condition for the determination of exchange rate changes. These shocks, motivated by the poor empirical performance of UIP equations, (see Sarno and Taylor, 2002 for an indicative survey) imply that the exchange rate does not move equiproportionately to interest rate differentials and in fact it often moves in the opposite direction.[12] These stochastic deviations, which can be thought of as excess returns in a particularly currency mean that the exchange rate can disconnect from the relative interests. The impulse responses show that a shock that brings about an initial exchange rate appreciation is equivalent to a demand shock as it depresses traded and non-traded wages via competition with overseas traded-sector wages. To deal with the temporary fall in wages, consumption - which is tilted up by the fall in domestic interest rates - is maintained by overseas borrowing and investment is stimulated by the fall in wages.

4.3. Variance decomposition

Table 2 shows the decomposition of unconditional variances for relative consumption, the real exchange rate and the current account from the model simulation. The first four columns show the contribution from each of the seven shocks in explaining the variance of these three key variables in the case of persistent, temporary UIP deviations and when the home economy is a creditor or debtor. The final three columns then exclude one type of shock in turn and shows the resulting contribution by the remaining shocks. Table 2 illustrates that

[12] In the appendix, A3, we outline how we estimate the stochastic deviations from UIP.

both sets of productivity shocks and UIP deviations are likely to play a dominate role in explaining the variance of the key open economy variables, the former for relative consumption and the latter for the real exchange rate and the current account.

The Table shows the dominant role that UIP deviations play under the baseline calibration in explaining the variance of the current account and real exchange rate over the business cycle. It also suggests that productivity shocks, particularly in the non-traded sector, might play an important role in explaining fluctuations in relative consumption and also for the real exchange rate and the current account when UIP deviations are excluded. Preference shocks play a negligible role in explaining the variances of these key variables unless we exclude productivity shocks altogether in which case they can explain over 20% of the variance in relative consumption. The finding that productivity shocks are important for quantities and relative prices even in the presence of exchange rate volatility is similar to other studies, such as Straub and Tchakarov (2004).

4.4. Simulated moments

We present second moments of the artificial simulated model in Table 3 for the benchmark calibration. The first column gives the moments from the UK data over the period 1980-2006. The next four columns of results correspond to the cases of persistent UIP deviations, temporary UIP and for the persistent UIP case also when the economy is a steady-state creditor or debtor - with assets or debts at 50% of GDP in each case, respectively. In the final three columns, we remove one set of shocks from the baseline calibration in order to understand how the artificial model data changes.[13]

The baseline calibration captures well the main moments of the data: consumption, labor inputs and wages are smooth relative to output and investment, the real exchange rate and the terms of trade are markedly volatile. The correlations of the main quantities and relative prices with output are all correctly signed (apart from interest rates). The model produces the positive relationship between the terms of trade and the real exchange rate found in the data and the exchange rate disconnect, with relative consumption negatively correlated. Finally, although higher than the data at 0.16, the model does not

[13] In earlier versions of this paper we also presented results for the estimated spill-over of productivity and preference shocks but as we found that these do not change the moments qualitatively we have removed them from this version.

predict that relative consumption will be perfectly correlated (with estimates in the range 0.5-0.7) and thus goes some distance towards understanding the lack of complete risk sharing.

This is because the non-state contingent bond is used to smooth investment and consumption following a shock.[14] In the event of a temporary productivity, which has little impact on permanent income. The home country consumer borrows from abroad, which raises overseas interest rates and lowers overseas consumption as well, which leads to a correlation in relative consumption. But when there are persistent productivity shocks, permanent income falls somewhat and so there is not as strong a need to borrow from abroad to smooth consumption or investment, which then means that overseas interest rates do not rise and lower overseas consumption. Hence there is something of a fall in the consumption correlation when there are non-state contingent bonds and persistent productivity shocks.

The persistence of the UIP shocks plays an important role in explaining both the relative variance of the real exchange rate and to a lessor extent that of relative consumption, which falls from 5.2% to 2.5% and from 1.1 to 0.9, respectively when we reduce the AR(1) persistence of UIP deviations from 0.88 to 0.38. Note also that the relative consumption becomes nearly acyclical (-0.02) when the UIP shocks fall in persistence. Moving towards a model where the steady-state level of net foreign assets is not zero does not alter the basic picture but when the home country is treated as a debtor investment, the real exchange rate and the terms of trade become more volatile and the current account becomes considerably less volatile.

If we examine the model with or without UIP deviations, compare column 2 to the final column, it appears that UIP deviations play a clear role in helping to explain the exchange rate disconnect. This is simply because the exchange rate can be driven whether there are movements in relative interest rates or not, which in turn depend mostly on planned relative consumption levels. An absence of UIP deviations from the model thus drives the correlation of relative consumption with real exchange rate to 0.76 rather than the data estimate of -0.61 or the benchmark model estimate of -0.65. Note also that in the model without UIP deviations, consumption, investment, labor inputs, real exchange rates and the terms of trade are somewhat too volatile. The main role of preference shocks it to raise the volatility of the labor input and lower that of the wage rate.

The overall performance of baseline calibrated model is reasonable. To

[14] Unlike an asset that can be bought to insure prior to shocks.

conclude the model performance in explaining the puzzles, we have (1) the model enables different shocks to interact and seems to solve the Backus-Smith puzzle and does not forecast perfect consumption correlation across the two economies with the help of a non-traded sector and incomplete financial markets; (2) this model stresses the HBS effect and therefore generates volatile real exchange rates; (3) countercyclical current account is a robust result, as the current account moves together with real exchange rate. In other words it seems to match the OECD and emerging economy experience suggested in Figure 1.

5. Model-data Comparison

A typical business cycle exercise examines the volatility of key economic variables and their correlation with output - as a measure of their business cycle behavior. At the very least such an examination neglects the cross-correlations in other elements in the VCM that may matter to us, which in this case is the relationship between exchange rates, relative consumption and the current account. Our model selection is thus based on the comparison of seven key variables of the VCM of endogenous variables simulated by our model to the actual data, see Appendix B for some further details. To illustrate our point, we consider the open economy sub-set of the variables for this exercise. In this section we obtain six statistical measures of distance of the model-generated data from the sample observations and the results are given in Table 4. The smaller statistics indicate a better fit of data to model and we find for the main model selection criterion the models with persistent UIP deviations with debtor status are closest to the observed data.

5.1. Model selection with VCM

If we choose to define a preferred model as that with the least deviation from the data, there may be a number of possible metrics we can employ. Our model selection from a class of candidate models is based on the comparison of the VCM of endogenous variables simulated by our model to that of the actual data, see Appendix B for further details on each test.[15] We consider a sub-set of the model variables that are closely related to the open economy puzzles highlighted in Figure 1: relative consumption, real exchange rate, relative output, home current account and home trade balance.

[15] A copy of the testing procedures written in MATLAB will be made available on request.

As well as basic criteria such as root mean squared error (RMSE) and mean absolute error (MAE), two likelihood ratio (LR) methods can be used to determine how different the two matrices are: (1) the Box-Bartlett test (1949 and 1937); (2) the distance measure flowing from the Kullback-Leibler (1951) Information Criteria (KLIC) method. We can also use the hypothesis testing method of Nagao (1973) and a revised test by Ledoit and Wolf (2002), which are designed to test an equality hypothesis of VCMs.[16] The key differences between these classes of approach are explained in the Appendix B but essentially the basic criteria of RMSE and MAE are akin to an approximate eyeballing of the data whereas the Box-Barlett test, KLIC methods, Nagao and Ledoit-Wolf allow for sampling variability and the KLIC also allows sampling variability in the simulated model.

For each case, we obtain six statistical measures of distance of the simulated model from the sample VCM for our 7 key variables. The results are given in Table 4. We assess the distance with different degrees of persistence in the UIP deviations and varying the NFA position. The smaller statistics indicate a better fit of data to the model. The best calibration according to each of the six criteria is therefore marked with an asterisk. If we examine the first three columns of results we will note that simple eyeballing of the data might lead us to prefer models with less persistent UIP shocks. But, when sampling and model uncertainty is accounted for, the other tests suggest we should prefer more persistence in the UIP deviations. We find models with persistent UIP deviations are closest to the observed data. Furthermore when we allow the steady-state debt position to move from creditor to debtor status we find that the best fit - smallest distance - occurs when the home economy is a debtor.

There are two main findings that stand out. Firstly, the distance measures suggest that persistent UIP deviations are helpful in generating a VCM similar to that of UK/US open economy data. We have shown in the impulse responses that deviations from UIP are the only forcing variable which helps resolves Backus-Smith puzzle, drives up large swings of real exchange rate and generates a volatile and countercyclical current account. More dominant UIP deviations are required to replicate the observed data. Secondly, we find that a non-zero NFA position

[16] The original Nagao's (1973) test is also an LR type test. The Ledoit and Wolf (2002) method aims to deal with the special cases where data dimension is larger than number of observations (or relatively small sample data). Such a property makes the data VCMs rank-deficient. Although we have rank-deficient VCMs in DSGE models for a different reason, where variables are greater in number than shocks and predetermined variables taken together, we utilize this method to deal with rank-deficiency problem. Note that canonical LR methods cannot be directly applied to rank-deficient VCMs. We outline our distance metrics in Appendix B.

is also helpful for improved goodness of fit. Where net debtor calibration for UK is slightly better than the net creditor case. However, negative or positive NFA position improve the model fit quite differently. A net debtor calibration mainly contributes to a better fit associated with current account dynamics. A net creditor calibration improves the goodness of fit for UK and US output and consumption data. In a two-country model, a net creditor UK means a net debtor US (as in real world). This realistic calibration can better explain relative output and consumption but also generate a volatile current account on both sides and thus create some distance for the overall fit. We therefore highlight a net creditor and persistent UIP deviation calibration for UK/US small open economy model according to the VCM distance approach.

5.2. Sensitivity analysis

The sensitivity analysis is shown in Figures 7-12 and is based on the seven-shock model with the basic calibration given in Table 1. We simulate the model and allow some deep parameters to change and check the sensitivity of some key moments with respect to several main statistical measures: the Backus-Smith correlation, the extent of exchange rate disconnect, the correlation between the trade and current account and the cyclicality of the current account. The vertical solid line(s) denotes the initial calibration.

First, we consider frictions in the model: costly investment and costly foreign asset holding. In Figure 7, although higher cost of investment alters volatility of open economy variables, it does not change the basic correlation structure. In Figure 8, costly foreign asset holding make the channel of risk sharing smaller, therefore the Backus-Smith correlation tend to zero. However, this will happen when the cost is extremely high. As the model has very simple assumption for financial markets, we emphasize its qualitative implication instead of its value denoted by basis points.

Secondly, we discuss the characteristics of the market and production. Steady state NFA does not alter real exchange rate dynamics significantly but it is crucial for current account dynamics. For a net debtor, a positive traded TFP shock leads to current account deficit. For example, upon a positive traded productivity shock, output increases, the real exchange rate appreciates, Home country borrows and a current account deficit results. But as a debtor there is requirement for paying interest, making the borrowing incentive lower and thus the extent to which the current account is countercyclical is mitigated, as shown in Figure 11.

Thirdly, we consider varying source of dynamics, the exogenous forcing variables. The UIP shock in the baseline calibration is highly persistent and by examining different degrees of persistence in UIP deviations as in Figure 10, we find there are real effects only in the case of highly persistent shocks. Adding UIP deviations reinforces the pattern of correlation we find in the data. When we vary the relative magnitude of non-traded productivity shocks in Figure 11, it leads to changes in the key correlations. A combination of relatively strong traded compared to non-traded productivity shocks contribute to negative Backus-Smith correlation and countercyclical current account. Turning to Figure 12, as preference shocks are strengthened, the negative correlation on both counts is weakened.

5.3. From model selection to parameter estimation

We can also replicate the sensitivity analysis for each of these key parameters but in terms of the distance measures rather than the base correlations in the data as in the previous section.[17] The diagnostics can be used to obtain estimates of the parameters that provide the best match to the data and essentially support the results outlined in the previous section.

It is clear that the minimum distance is achieved when treating the cost of investment is in the neighborhood of 2. The required costs of financial intermediation (ϵ) do not seem to have to be especially high when we examine the sensitivity analysis on cost of financial intermediation, i.e., the spread between return on foreign and domestic bond. This parameter affects the trade-off between home and foreign bonds. But the four criterion all suggest model will improve when the spread increases somewhat. We attribute this result to strong home bias in asset holding. The adoption of the assumption that the home economy is a net debtor also seems to help model fit.[18]

Finally we examine the correct level of persistence for the shock processes. More persistent UIP deviations are, in general, preferred and increasing the relative volatility of traded to non-traded shocks seems to help the fit. But as with earlier results (see Chadha *et al*, 2001) increasing the volatility of preference

[17] These Figures are excluded from this version of the paper in the interest of space but are available on request.

[18] Although the UK has a steady-state level of debt near zero and Lane and Milesi-Ferretti (2002) document the mean net foreign asset position to GDP at 6% i.e., UK is a small net creditor. Our approach seems to locate the correct approximate region for the level of steady-state debt i.e. not very far from zero.

shocks does not seem to improve the fit of the model markedly.

6. Conclusion

Open-economy general equilibrium models offer an attractive laboratory in which to examine the insolubility or otherwise of data puzzles. We examine the properties of a two-sector real business cycle model with incomplete financial markets. The model is driven by a number of driving forces in both domestic and overseas traded and non-traded productivity, to the work-leisure margin at home and overseas and to deviations in the exchange rate from the level suggested by the UIP equation.

We find some evidence to support the proposition that when all these shocks perturbate the model economy there is some move towards resolution of the puzzles. The most important modelling choices - over and above a standard one-sector small economy RBC model - involve the adoption of a two sector model, allowing for shocks to non-traded as well as traded sector productivity, the employment of incomplete markets with the existence of a non-state contingent bond and of stochastic deviations from the UIP equation for the exchange rate. The aspects of the model induce greater real exchange rate variability and yet alongside the absence of complete risk sharing ensure that consumption need not simultaneously jump to arbitrage price differentials.

Finally we note that the modelling approach we use is flexible enough to allow examination of deep parameters for small open economies. And for the researcher to examine some simple summary statistics when assessing model fit. These measures might usefully be applied more generally to the question of the fit of data to DSGE models.

Appendix

A. Measurement of Exogenous Shocks

A.1. Productivity Measurement

Sectoral productivity is calculated as total factor productivity (TFP) in traded (manufacturing) or non-traded sector (services). We use OECD STAN database 2005 release to construct sectoral TFP series for UK and US. Incomplete data on total hours and gross capital stock is complemented by total employment data and capital formation data.

$$TFP_t^{T,NT} = \log\left(\frac{Y_t^{T,TN}}{\left(K_t^{T,TN}\right)^{1-\alpha}\left(N_t^{T,TN}\right)^{\alpha}}\right) \tag{A.1}$$

A.2. Measuring the preference shock

Preference measures, ξ_t, are calculated from the Euler equation for leisure-labor output and solved for ξ_t with output, total hours, wage or consumption. The time endowment and the utility non-separable to leisure are:

$$L_t = 1 - l_t \tag{A.2}$$

$$U = \frac{1}{1-\rho} C_t^{1-\rho} \frac{L_t^{\eta\xi_t}}{\eta\bar{\xi}} \tag{A.3}$$

In above utility function, the shock ξ_t is specified to be leisure-biased. We make the percentage deviation $\widehat{\xi}_t = -1\%$ to see the impulse to a preference shock biased to consumption.

$$w_t = \frac{U_L}{U_C} = \frac{\eta}{(1-\rho)} \frac{C_t \xi_t}{(1-l_t)} \tag{A.4}$$

Preference measure is the detrended series of ξ_t in logarithm:

$$\xi_t = \ln\left(\frac{(1-\rho)}{\eta} \frac{w_t(1-l_t)}{C_t}\right) \tag{A.5}$$

The preference shock is measured by calculating ξ_t with US and UK aggregate data.

A.3. Stochastic deviations from UIP

We allow for deviations in the UIP condition for the exchange rate, making exchange rate volatility attributable to more factors. Our version of UIP shocks is a simple treatment allowing market participants in foreign exchange markets to let the exchange rate deviate from theoretical value in the short run. The nominal exchange rate adjustment is according to UIP and a shock $x_{u,t}$ by:

$$E_t \Delta s_{t+1} = i_t - i_t^* + \varepsilon \widehat{B}_t + x_{u,t} \tag{A.6}$$

From Selaive and Tuesta's (2003) estimation on US data, we take the calibration for ε. Kollmann (2003) uses a two-part UIP shock $x_{u,t} = a_t + \omega_t$ and find UIP shock is quite persistent. We compare his calibration with a temporary UIP shock scenario.

B. Testing Model Fit

Canova and Ortega (2000) discuss four possible approaches in evaluating DSGE model fit. The variety of approached arises from the different treatment of model uncertainties and data sampling uncertainties: (a) an informal approach, which ignores both sampling variability in the data and uncertainty regarding model parameters, (b) methods that consider model uncertainty but not sampling variability in the data, and (c) methods that consider sampling variability in the data but not uncertainty in model; and (d) approaches that account for both sampling variability in the data and model uncertainty.

As in Bhattacharjee and Thoenissen (2007), they use the modified Nagao test which belongs to the class of method (c):

> "... we consider an approach that uses sampling variability of actual data to provide a measure of the distance between model and the data, holding the model VCM fixed. This approach is explicitly based on the context of dynamic general equilibrium macroeconomic models, where given specific calibrated or estimated values for the parameters, the model can be simulated for as many periods of time as desired. Thus, for given parameter values, the asymptotic VCM of the state variables obtained from such simulation has no sampling variability. On the other hand, the data VCM is based on a data for a finite sample period. In most applications, this period would be from 1960 or later to the most recent period for which data are available. Thus, there is substantial sampling variation in the data VCM, while the model VCM can be considered fixed for a given combination of parameter values. By computing distances for distinct combinations of possible parameter values across all the competing models, we can ignore the uncertainty regarding calibration or estimation of parameters, while taking account of sampling variability in the actual data."

Similarly Box-Bartlett and Ledoit-Wolf are alternative methods derived from approach (c). In addition we explore the possibilities of using parallel approaches following Canova and Ortega's (2000) guideline: eyeballing approach such as RMSE and MAE are implementations of approach (a); Kullback-Liebler is an implementation of approach (d).

Since most DSGE models are driven by only a limited number of shocks and predetermined state variables, the model VCM is usually rank-deficient. Except for RMSE and MAE, we use a projection of both data and model VCM to lower dimensional subspace introduced by Bhattacharjee and Thoenissen (2007) to deal with the rank-deficient problem.

The methods developed here will also take into account two other common features of model selection in the stated context. First, as emphasized earlier, DSGE models are intended to be abstractions of reality and are often driven by a lesser number of shocks than the number of state variables. In other words, while actual data VCMs would be full-rank, simulated data VCMs may often have a lower rank. Our methods will explicitly take into account this possibility. Second, the metrics will be developed in such a way that enables model selection when the candidate DSGE models may be non-nested. This feature of our methodology will also obviously important and enhance the applicability of the methods.

B.1. Distance metrics

We denote by $[\Sigma_0]_{m\times m}$ the full-rank data VCM estimated using n_0 data points ($\rho(\Sigma_0) = m$), where ρ is the rank of VCM. $[\Sigma_{M_1}]_{m\times m}, [\Sigma_{M_2}]_{m\times m}, [\Sigma_{M_3}]_{m\times m}, \ldots$ denote estimated VCMs using simulated data from a countable collection of competing models $M_1, M_2, M_3, \ldots$ and based on $n_1, n_2, n_3, \ldots$ simulated observations respectively. Some of these matrices may be rank deficient ($\rho(\Sigma_{M_j}) \leq \rho(\Sigma_0) = m$).

We shall propose several alternate metrics, denoted $d(\Sigma_0, \Sigma_{M_j})$, that give scalar measures of how different any of the simulated VCMs are from Σ_0, where d is a metric measuring the distance between Σ_0 and Σ_{M_j}. These measures can then be used select an appropriate model from all the competing ones. In the following, we focus on one competing model VCM, say Σ_M and elaborate on different possible approaches and corresponding metrics.

B.1.1. Naive, or Eyeballing, approach

This is not based on any distributional assumption. Root Mean Squared Errors (RMSE) and Mean Absolute Errors (MAE) are defined as:

$$RMSE = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \widetilde{\sigma}_{i,j}^2$$

$$MAE = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} |\widetilde{\sigma}_{i,j}|$$

where $\widetilde{\Sigma} = ((\widetilde{\sigma}_{ij}))_{m \times m} = \Sigma_M - \Sigma_0$. In terms of the typology developed in Canova and Ortega (2000), the above two metrics ignore sampling variability in both data and model VCM.

B.1.2. Testing approach

This approach is based on a multivariate normality assumption underlying both the estimated VCMs, Σ_0 and Σ_M. However, we consider the possibility that the model VCM may not be full rank. The idea here is to pretend that we are conducting a test of the hypothesis $H_0 : \Sigma_0 = \Sigma_M$ against the omnibus alternative $H_1 : \Sigma_0 \neq \Sigma_M$. We are not as such interested in the outcome of the test, since we do not strongly believe that any of the models will generate simulated VCMs that are statistically indistinguishable from the data VCM. However, we can still use the p-values of the tests (or the values of the test statistic itself, adjusted for degrees of freedom) to give us a metric to compare between competing models. Note that the testing approach considers sampling variation in the data VCM, but the comparison is made with a simulated VCM based on large data where sampling variability may be negligible. We consider the following cases:

Σ_M is full-rank Here we can use a whole battery of tests developed in the multivariate statistics literature. The most popular of these tests are the Box (1949) modification to the test proposed by Bartlett (1937), and the test proposed by Nagao (1973).

Bartlett (1937) proposed the test statistic:

$$M = \sum (n_0 + n_M) \ln |\Sigma| - n_0 \ln |\Sigma_0| - n_M \ln |\Sigma_M|$$

where the pooled estimate of the common covariance matrix under the null hypothesis is

$$\Sigma = \frac{1}{n_0 + n_M}\left[n_0\Sigma_0 + n_M\Sigma_M\right].$$

When multiplied by a scaler C^{-1} (Box, 1949):

$$C^{-1} = 1 - \frac{2m^2 + 3m - 1}{6\,(m+1)}\left(\frac{1}{n_0} + \frac{1}{n_M} - \frac{1}{n_0 + n_M}\right),$$

the Box's M test statistic MC^{-1} has a Chi-square distribution (df$= m\,(m+1)\,/2$) under the null hypothesis and multivariate normality assumption.

Nagao (1973) proposed a test for the null hypothesis $H_0 : \Sigma_M^* = I$ against the omnibus alternative (where I is the identity matrix) given by the test statistic:

$$N = \frac{n_M}{2} tr\left(\Sigma_M^* - I\right)^2,$$

where $tr(.)$ denotes trace of a square matrix. The test statistic has a Chi-square distribution (df$= m\,(m+1)\,/2$) under the null hypothesis and multivariate normality assumption. This test can be adopted to our situation by using the Cholesky decomposition of Σ_0, as follows:

$$\begin{aligned} \Sigma_0 &= P'P \\ \Sigma_M^* &= P'^{-1}\Sigma_M P^{-1} \\ I &= P'^{-1}\Sigma_0 P^{-1} \end{aligned}$$

so that testing $H_0 : \Sigma_0 = \Sigma_M$ is now equivalent to testing $H_0 : \Sigma_M^* = I$ against the omnibus alternative. This is equivalent to premultiplying the actual and simulated data vectors by P'^{-1}. Both the Box's M-test and Nagao's test are known to be very conservative even in small samples (seldom accept the null hypothesis); this is, however, not of any major consequence for our work since we are not interested in the exact results of the test.

Σ_M is rank deficient ($\rho(\Sigma_M) < \rho(\Sigma_0) = m$) This is the usual case. The model here is clearly an abstraction driven by only a limited number of shocks. In fact, this abstraction can also represent reality to a high degree, in the sense that often only a small number of shocks can explain a substantial part of the variation in actual data on a larger number of state variables. In most applications, only a limited number of leading eigenvalues (and their corresponding eigenvectors) account for most of the variation in the data VCM, the remaining eigenvalues are small in comparison.

While the Box-Bartlett and Nagao tests do not directly apply to this situation, we propose two simple modifications. First, we adapt an extension of Nagao's test to the rank deficient case proposed by Ledoit and Wolf (2002). Ledoit and Wolf (2002) have recently considered a situation where the number of variables is large and higher than the sample size. They modify the Nagao (1973) test to this situation and derive asymptotic theory when both the dimension of the VCM and sample size increase to ∞ at the same asymptotic rate. In particular, their test statistic is given by:

$$W = \frac{1}{m} tr\left(\Sigma_M^* - I\right)^2 - \frac{m}{\rho(\Sigma_M^*)}\left[\frac{1}{m} tr\left(\Sigma_M^*\right)\right]^2 + \frac{m}{\rho(\Sigma_M^*)}.$$

Under the null hypothesis and multivariate normality, $\frac{1}{2}\rho(\Sigma_M^*).m.W$ has a Chi-squared distribution with $m(m+1)/2$ degrees of freedom. This extension is based on an asymptotic setup where, as sample size (time periods under study) increases, the set of state variables under comparison is also augmented; this assumption is reasonable in many practical situations.

Second, following Bhattacharjee and Thoenissen (2007), we project the data VCM onto a lower dimensional subspace spanned by the shocks and free predetermined variables driving the model. The usual Box-Bartlett and Nagao tests are then employed for VCM comparisons over this lower dimensional subspace; see Bhattacharjee and Thoenissen (2007) for further details.

B.1.3. Measures based on distance between distributions

One possible limitation of the above testing based approach is that it ignores sampling variation in the model VCM, and therefore its applicability for moment comparison specific to known time periods may be tenuous. An alternative is the approach, indicated in Watson (1993), based on computing the Kullback-Liebler Information Criteria (KLIC) between the distributions given by the data (mean zero, VCM Σ_0) and the model (mean zero, VCM Σ_M) and choosing the best model based on this measure. The KLIC is given by:

$$I\left(\Sigma_0, \Sigma_M\right) = E_{f(.;0,\Sigma_0)} \ln \frac{f(Y;0,\Sigma_M)}{f(Y;0,\Sigma_0)} = \int_{-\infty}^{\infty} \ln \frac{f(y;0,\Sigma_M)}{f(y;0,\Sigma_0)} f(y;0,\Sigma_0) dy,$$

where $f(.;0;\Sigma)$ denotes the density of the multivariate Gaussian distribution with mean vector zero and VCM Σ, and the expectation is taken with respect to the distribution of the data (mean zero and VCM Σ_0).

While Watson (1993) suggests use of the KLIC in full-rank situations, we extend the method to models with lower number of shocks by using density functions for singular normal distributions. Specifically, we consider the singular value decomposition (SVD) of the simulated model VCM : $\Sigma_M = \lambda_1 e_1 e_1' + \lambda_2 e_2 e_2' + ... + \lambda_p e_p e_p' + 0 e_{p+1} e_{p+1}' + ... + \lambda_m e_m e_m'$, where $p = \rho(\Sigma_M) < m$ is the rank of the model VCM. The density function of this rank-deficient model (mean zero, VCM Σ_M, $\rho(\Sigma_M) = p < \rho(\Sigma_0) = m$) on the subspace spanned by only the p leading eigenvectors is:

$$f\left(\underline{y}_{m\times 1}; 0, \Sigma_M\right) = \frac{1}{(2\pi.\lambda_1.\lambda_2....\lambda_p)^{m/2}} . \exp\left(-\frac{1}{2}\underline{y}'\Sigma_M^{-}\underline{y}\right),$$

where the generalized inverse (g-inverse) of Σ_M is given by $\Sigma_M^{-} = 1/\lambda_1.\underline{e_1}.\underline{e_1'} + 1/\lambda_2.\underline{e_2}.\underline{e_2'} + \ldots + 1/\lambda_p.\underline{e_p}.\underline{e_p'}$. The density of the data VCM (full-rank) is computed in the usual way.

The KLIC approach, however, has a few features that are of importance. First and most importantly, KLIC does not give a strict distance metric, since it is not symmetric in its arguments. One can use symmetric versions of KLIC reported in the literature and besides this may not be a major issue in our case, since we are interested only in finding distances of different models from the data VCM, and to that extent our approach is consistent. Second, KLIC is of course based on an assumed parametric distribution. We may assume multivariate normality or if appropriate, some other parametric distribution. Third, the KLIC is often difficult to compute particularly in a multi-dimensional case because this involves numerical integration in high dimensions. We bypass this problem by taking a Monte Carlo or bootstrap approach as follows.

We note that the KLIC is the expected value of difference of log-likelihoods under the two alternative distributions (given by Σ_0 and Σ_M) for samples from the distribution given by the data VCM. Empirically we can either generate a Monte Carlo sample (sample size N_0^{MC}) with data VCM, or take bootstrap resamples (bootstrap sample size N_0^{BS}) from the actual data, and then calculate the sample mean of log likelihood ratios. By the weak law of large numbers, both these approaches will give consistent estimates of the KLIC. However, the Monte Carlo method will depend more specifically on the validity of the multivariate normality assumption, hence the bootstrap approach may be preferable in practice:

$$\widehat{I}_{Monte\ Carlo}\left(\Sigma_0, \Sigma_M\right) = \frac{1}{N_0^{MC}} \sum_{i=1}^{N_0^{MC}} \ln \frac{f(y_i; 0, \Sigma_M)}{f(y_i; 0, \Sigma_0)},$$

$$\widehat{I}_{Bootstrap}\left(\Sigma_0, \Sigma_M\right) = \frac{1}{N_0^{BS}} \sum_{i=1}^{N_0^{BS}} \ln \frac{f(y_i; 0, \Sigma_M)}{f(y_i; 0, \Sigma_0)}.$$

MATLAB codes for the implementation of the metrics used in this paper are available from the authors on request.

References

[1] Backus, David K., Patrick J. Kehoe, and Finn E. Kydland, (1995). "International Business Cycles: Theory and Evidence", in Thomas F. Cooley (ed.) *Frontiers of Business Cycle Research*, Princeton University Press, Princeton, pp.331-56.

[2] Backus, David K. and Gregor W. Smith, (1993). "Consumption and real exchange rates in dynamic economies with non-traded goods", *Journal of International Economics*, Elsevier, vol. 35(3-4), pp.297-316.

[3] Bartlett, Maurice, (1937). "Properties of sufficiency and statistical tests", *Proceedings of the Royal Society of London Series A* 160, pp.268–282.

[4] Baxter, Marianne and Mario J. Crucini, (1995). "Business Cycles and the Asset Structure of Foreign Trade", *International Economic Review*, vol. 36(4), pp.821-54.

[5] Baxter, Marianne and Urban J. Jermann, (1997). "The International Diversification Puzzle Is Worse Than You Think", *American Economic Review*, vol. 87(1), pp.170-80.

[6] Benigno, Pierpoalo. (2001). "Price stability with imperfect financial integration", New York University, *mimeo.*

[7] Benigno, Gianluca and Christoph Thoenissen, (2004). "Consumption and Real Exchange Rates with Incomplete Markets and Non-traded Goods", *CDMA Conference Paper Series 0405*, Centre for Dynamic Macroeconomic Analysis. Forthcoming in *Journal of International Money and Finance.*

[8] Bhattacharjee, Arnab and Christoph Thoenissen, (2007). "Money And Monetary Policy In Dynamic Stochastic General Equilibrium Models," *Manchester School*, vol. 75(s1), pp.88-122, 09.

[9] Box, George E. P., (1949) "A general distribution theory for a class of likelihood criteri", *Biometrika*, vol. 36, pp.317–346.

[10] Canova, Fabio and Ortega, Eva, (2000). "Testing Calibrated General Equilibrium Mod-els", in Mariano, R., T. Shuermann and M. Weeks (eds.) *Inference using Simulation Techniques*, Cambridge University Press

[11] Chadha, Jagjit. S, Norbert Janssen and Charles Nolan, (2001). "Productivity and preferences in a small open economy", 2001, *Manchester School,* vol. 69, pp.57-80.

[12] Chari, Varadarajan V., Patrick J. Kehoe, and Ellen R. McGrattan, (2002). "Can sticky price models generates volatile and persistent real exchange rate?", *Review of Economic Studies*, vol. 69, pp.633-63.

[13] Corsetti, Giancarlo, Luca Dedola and Sylvain Leduc, (2004). "International risk-sharing and the transmission of productivity shocks", *ECB Working Paper Series* No.308. Forthcoming in the *Review of Economic Studies.*

[14] den Haan, Wouter, (2000). "The comovement between real activities and price at different business cycle frequencies", *Journal of Monetary Economics*, vol. 46, pp.3-30.

[15] Devereux, Michael B. and Charles Engel, (2002). "Exchange rate pass-through, exchange rate volatility, and exchange rate disconnect", *Journal of Monetary Economics*, vol. 49(5), pp.913-940.

[16] Dornbusch, Rudiger, (1976). "Expectations and Exchange Rate Dynamics", *Journal of Political Economy*, vol. 84, pp.1161-76.

[17] Engel, Charles, (2000). "Comment on Obstfeld and Rogoff", *NBER Macroeconomics Annual.* National Bureau of Economic Research.

[18] Fitzgerald, Doireann, (2003). “Terms-of-trade effects, interdependence and cross-country differences in price levels”, manuscript, UC-Santa Cruz.

[19] Frankel, Jeffrey, (1996). “How Well Do Foreign Exchange Markets Work: Might a Tobin Tax Help?” in Mahbub ul Haq, Inge Kaul, and Isabelle Grunberg, eds., The Tobin Tax: Coping with Financial Volatility. New York and Oxford: Oxford University Press.

[20] Geweke, John, (1999). “Computational Experiments and Reality”, *mimeo.*

[21] Hall, Robert. E., (1997). “Macroeconomic fluctuations and the allocation of time”, *Journal of labor Economics*, vol. 15, pp.223-228.

[22] Kakkar, Vikas, (2003). “The relative price of non-traded goods and sectoral total factor productivity: an empirical investigation”, *Review of Economics and Statistics*, vol. 85(2), pp.444-452.

[23] Kehoe, Patrick J. and Fabrizio Perri, (2002). “International business cycles with endogenous incomplete markets”, *Econometrica*, vol. 70 (3) pp.907-928.

[24] King, Robert G. and Mark W. Watson, (1998). “The solution of singular linear difference systems under rational expectations”, *International Economic Review*, vol. 39(4), pp.1015-26.

[25] Kollmann, Robert, (2003). “Monetary Policy Rules in an Interdependent World”, CEPR Discussion Paper No. 4012.

[26] Kullback, Solomon, and Leibler, Richard A., (1951). “On information and sufficiency”, *Annals of Mathematical Statistics*, vol. 22. pp.79-86.

[27] Lane, Philip R. and Gian Maria Milesi-Ferretti, (2002). “External wealth, the trade balance, and the real exchange rate”, *European Economic Review*, vol. 46(6), pp.1049-1071.

[28] Ledoit, Oliver and Michael Wolf, (2002). “Some hypothesis tests for the covariance matrix when the dimension is large compared to the sample size", *Annals of Statistics* 30, pp.1081–1102.

[29] Nagao, Hisao, (1973). "On some test criteria for covariance matrix". *Annals of Statistics*, vol. 1, pp.700-709.

[30] Obstfeld, Maurice and Kenneth Rogoff, (2000). "The Six Major Puzzles in International Macroeconomics: Is There a Common Cause?," *NBER Macroeconomics Annual.* National Bureau of Economic Research.

[31] OECD, (2004). *STAN database.*

[32] Sarno, Lucio and Mark P. Taylor, (2002). *The Economics of Exchange Rates*, Cambridge University Press.

[33] Schmitt-Grohe, Stephanie and Martin Uribe, (2003). "Closing small open economy models", *Journal of International Economics,* vol. 61(1), pp.163-185.

[34] Selaive, Jorge D. and Tuesta, Vincente (2003). "Net foreign assets and imperfect pass-through: the consumption-real exchange rate anomaly". Board of Governors of the Federal Reserve System, International Finance Discussion Papers, No. 764.

[35] Sorensen, Bent E. & Wu, Yi-Tsung & Yosha, Oved & Zhu, Yu, (2007). "Home bias and international risk sharing: Twin puzzles separated at birth", *Journal of International Money and Finance*, vol. 26(4), pp.587-605.

[36] Stockman, Alan C and Linda L. Tesar, (1995). "Tastes and Technology in a Two-Country Model of the Business Cycle: Explaining International Comovements", *American Economic Review*, vol. 85(1), pp.168-85.

[37] Straub, Roland and Ivan Tchakarov (2004). "Non-fundamental exchange rate volatility and welfare", *ECB Working Paper Series* No. 328.

[38] Tesar, Linda L., (1993). "International Risk Sharing and Non-Traded Goods", *Journal of International Economics*, vol. 35 (1-2), pp.69-89.

[39] Tesar, Linda L. and Ingrid M. Werner, (1995). "Home bias and high turnover", *Journal of International Money and Finance*, vol. 14(4), pp.467-492.

[40] Watson, Mark, W., (1993). "Measures of Fit for Calibrated Models", *Journal of Political Economy*, vol. 101(6), pp.1011-1041.

Table 1 - Quarterly Calibration for Small Open Economy Model

Parameter	Values	Description
β	0.99	Discount factor
δ	0.025	Depreciation factor
α	0.67	labor share
ρ	2	CRRA
$\eta\overline{\xi}$	-4	Elasticity of marginal value of time
θ	1.5	Elasticity: Home/Foreign traded goods
κ	0.74	Elasticity: Traded/Non-traded goods
(υ, υ^*)	$(0.73, 0.02)$	Home produced share of tradeables in home and overseas
(ω, ω^*)	$(0.45, 0.45)$	Share of tradeables in home and overseas output
ε	70 basis points	Interest spread (quarterly)
$\overline{a}$	0	Steady state Net Foreign Asset
b	10	Cost of capital adjustment
(ρ_A, ρ_{A^*})	0.918	Persistence of traded productivity shocks
(σ_A, σ_{A^*})	$(1.17\%, 1.41\%)$	Volatility of traded productivity shocks
(ρ_{AN}, ρ_{AN^*})	0.945	Persistence of non-traded productivity shocks
$(\sigma_{AN}, \sigma_{AN^*})$	$(0.51\%, 0.56\%)$	Volatility of non-traded productivity shocks
(ρ_ξ, ρ_{ξ^*})	0.937	Persistence of preference shocks
$(\sigma_\xi, \sigma_{\xi^*})$	$(0.82\%, 0.82\%)$	Volatility of preference shocks
$(\rho_{UIP^H}, \rho_{UIP^L})$	0.88 or 0.38	Persistence of UIP deviations (high or low)

Note: We have an utility function similar to Chadha *et al* (2001). The elasticity of intertemporal substitution in leisure $\frac{1}{\eta\overline{\xi}-1}$ is -0.2; the elasticity of labor supply in this model is around 4; the discount factor β, CRRA ρ, depreciation coefficient δ and labor share α are taken from standard open economy and real business cycle literature such as Corsetti *et al* (2005), Chari *et al* (2002); we take elasticity of substitution among consumables θ, κ from Corsetti *et al*; the share of traded goods ω, ω^* are taken as 0.45 in accordance with the literature; for home bias feature in traded goods, we take average value share of UK produced goods in UK and US GDP, υ and υ^*, respectively; interest spread ε is a yield discount when holding foreign bond and is calibrated as 280 base points annually by Selaive and Tuesta (2003); the cost of capital adjustment b is calibrated to match UK output volatility; we set Net Foreign Asset position $\overline{a}$ as zero in benchmark case; the persistence and volatility of shocks are estimated on UK data.

Table 2 - Variance Decomposition of Current Account, Real Exchange Rate and Relative Consumption

	(a) All shocks				(b) Excluding shocks:		
	Persistent UIP Dev.	Temporary UIP Dev.	Net Creditor[1]	Net Debtor[1]	Productivity	Preference	UIP
Current Account (CA/Y)							
Traded productivity, Home	0.2%	1.6%	0.1%	0.4%	–	0.2%	28.6%
Non-Traded productivity, Home	0.4%	3.1%	1.0%	3.1%	–	0.4%	55.6%
Traded productivity, Foreign	0.0%	0.1%	0.0%	0.0%	–	0.0%	1.5%
Non-Traded productivity, Foreign	0.0%	0.0%	0.0%	0.0%	–	0.0%	0.1%
Preference, Home	0.1%	0.5%	0.1%	0.1%	0.1%	–	9.0%
Preference, Foreign	0.0%	0.3%	0.0%	0.1%	0.0%	–	5.2%
UIP	99.3%	94.4%	98.8%	96.3%	99.9%	99.4%	–
Real Exchange Rate (RER)							
Traded productivity, Home	0.8%	5.6%	0.9%	0.8%	–	0.8%	35.9%
Non-Traded productivity, Home	0.0%	0.1%	0.0%	0.0%	–	0.0%	0.7%
Traded productivity, Foreign	0.1%	0.6%	0.1%	0.1%	–	0.1%	3.8%
Non-Traded productivity, Foreign	0.5%	3.2%	0.5%	0.4%	–	0.5%	20.1%
Preference, Home	0.4%	2.7%	0.4%	0.4%	0.4%	–	17.4%
Preference, Foreign	0.5%	3.5%	0.5%	0.5%	0.5%	–	5.2%
UIP	97.7%	84.3%	97.6%	97.8%	99.1%	98.6%	–
Relative Consumption (CC^*)							
Traded productivity, Home	14.4%	16.0%	14.5%	14.3%	–	15.8%	16.2%
Non-Traded productivity, Home	62.8%	70.0%	62.6%	63.0%	–	69.2%	71.0%
Traded productivity, Foreign	0.8%	0.9%	0.8%	0.8%	–	0.9%	0.9%
Non-Traded productivity, Foreign	1.3%	1.4%	1.3%	1.3%	–	1.4%	1.4%
Preference, Home	4.7%	5.3%	4.8%	4.7%	22.8%	–	5.3%
Preference, Foreign	4.5%	5.0%	4.5%	4.5%	21.9%	–	5.1%
UIP	11.5%	1.3%	11.6%	11.4%	55.4%	12.6%	–

Table 3 - Results of benchmark UK/US calibration

	UK Data	(a) All shocks				(b) Excluding shocks:		
		Persistent UIP Dev.	Temporary UIP Dev.	Net Creditor[1]	Net Debtor[1]	Productivity	Preference	UIP
Relative volatility to output (interest rate and CA/Y take raw value)								
Consumption	*0.78*	0.92	0.64	0.91	0.93	1.09	1.04	0.59
Investment	*2.30*	3.99	2.84	3.91	4.08	4.62	4.58	2.60
Interest rate	*1.01*	0.42	0.37	0.40	0.43	0.42	0.42	0.05
Labour	*0.95*	0.90	1.00	0.89	0.90	1.13	0.36	1.02
Wage	*0.86*	0.89	0.65	0.88	0.90	1.01	1.02	0.61
RER	*4.89*	5.16	1.91	5.13	5.20	6.71	6.21	0.67
ToT	*1.66*	5.88	2.47	5.85	5.91	7.50	7.08	1.46
CA/Y	*1.06*	1.99	0.61	4.12	2.80	1.99	1.99	0.12
CC*	*1.27*	1.13	0.94	1.12	1.14	1.22	1.24	0.92
Correlation with output								
Consumption	*0.79*	0.90	0.97	0.90	0.89	0.91	0.90	1.00
Investment	*0.79*	0.88	0.93	0.89	0.88	0.89	0.90	0.96
Interest rate	*0.19*	−0.33	−0.08	−0.33	−0.33	−0.51	−0.43	0.90
Labour	*0.78*	0.56	0.79	0.57	0.56	0.55	0.14	0.82
Wage	*0.13*	0.56	0.32	0.56	0.55	0.37	0.94	0.27
RER	*-0.10*	−0.46	0.00	−0.46	−0.45	−0.63	−0.58	0.57
ToT	*-0.13*	−0.47	−0.05	−0.47	−0.46	−0.64	−0.59	0.19
CA/Y	*-0.30*	−0.52	−0.22	−0.36	−0.22	−0.65	−0.64	−0.13
CC*	*0.19*	0.64	0.48	0.65	0.64	0.83	0.65	0.46
Correlation with RER								
ToT	*0.10*	0.98	0.81	0.98	0.98	1.00	0.98	0.21
CC*	*-0.61*	−0.65	−0.02	−0.65	−0.66	−0.83	−0.75	0.76

Table 4 - Model Selection by Variance Covariance Matrix (VCM) Distance

VCM Distance Calculation Method	All Shocks Persistent UIP Dev.	Temporary UIP Dev.	i.i.d UIP Dev.	Net Creditor	Net Debtor
RMSE	0.0462%	0.0431%	0.0444%	0.0620%	0.0388%∗
MAE	0.0241%	0.0122%	0.0121%∗	0.0313%	0.0204%
Box-Bartlett	92053	99787	99841	44379	43947∗
Kullback-Leibler	432	468	469	208	206∗
Nagao	2.74×10^7	3.18×10^7	3.18×10^7	5.12×10^6	4.95×10^6∗
Ledoit-Wolf	2.70×10^7	3.14×10^7	3.14×10^7	5.02×10^6	4.85×10^6∗

Note to Tables 2 to 4: The quarterly data is the HP filtered series of OECD MEI, 1980-2006: RER is real exchange rate; CA/Y is current account to GDP ratio; ToT is terms of trade and is import price over export price; CC* is relative consumption to US. The basecase calibration is as Table 1. The UK (home country) and net creditor calibration denotes $\overline{a} = 0.5$ while net debtor denotes $\overline{a} = -0.5$ for a small open economy; the persistent UIP shock denotes an AR(1) persistence coefficient of $\rho_{UIP} = 0.88$ as in Kollmann (2003); whereas the temporary case and i.i.d case take the value of $\rho_{UIP} = 0.38$ and $\rho_{UIP} = 0$, respectively; in both net creditor and net debtor case the UIP deviations are persistent; RMSE denotes root mean squared errors; MAE denotes mean squared errors; each of the calculation methods is discussed in Appendix B; the asterisk (*) denotes minimum distance measure in all five calibration.

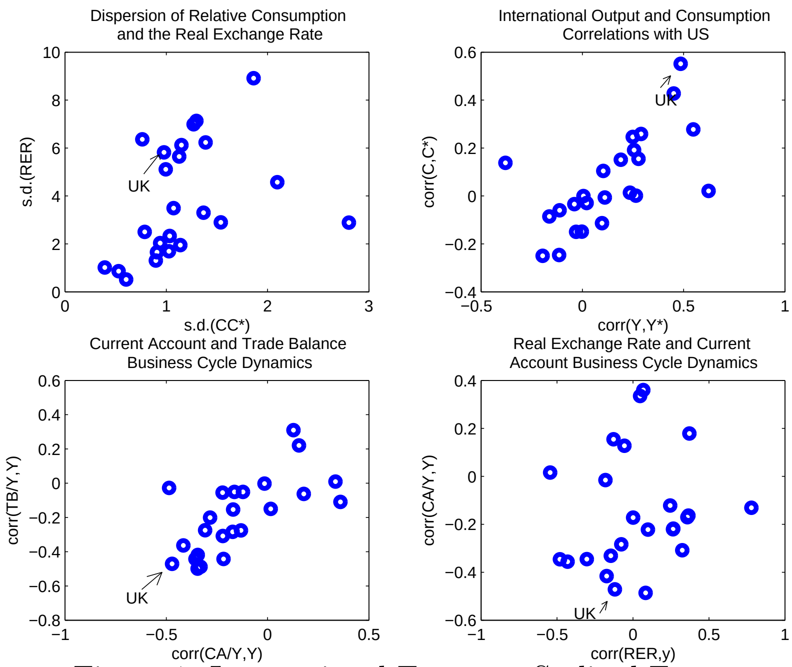
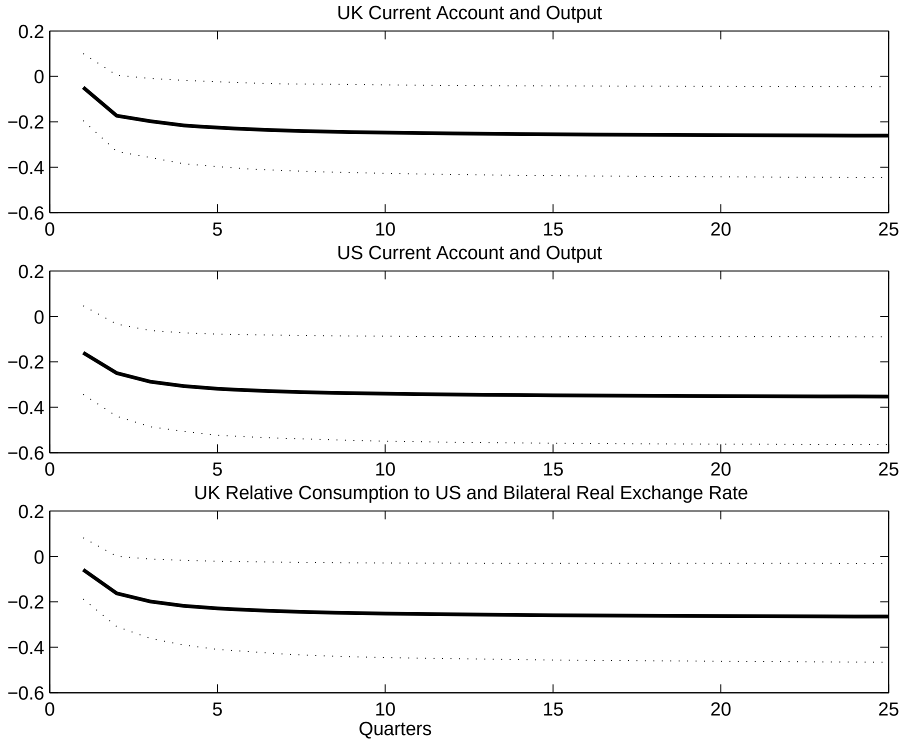


Figure 1: International Economy Stylised Facts

Note: Quarterly data of 24 OECD and emerging market economies is obtained from IMF IFS database. s.d. denotes standard deviation of HP-filtered series of the variables. corr denotes the correlation coefficient between two HP-filtered series. RER denotes bilateral real exchange rate. C, C*, Y, Y* are household consumption and real GDP of small open economy and US. CC* is the relative consumption to US. TB/Y is the ratio of trade balance to output and CA/Y is the ratio of current account to output.



Note: Forecast error correlation calculated with Den Haan (1996) code. Solid lines denote correlations and the dotted lines represent a 90% confidence interval of point estimate.

Figure 2: Price Stickiness

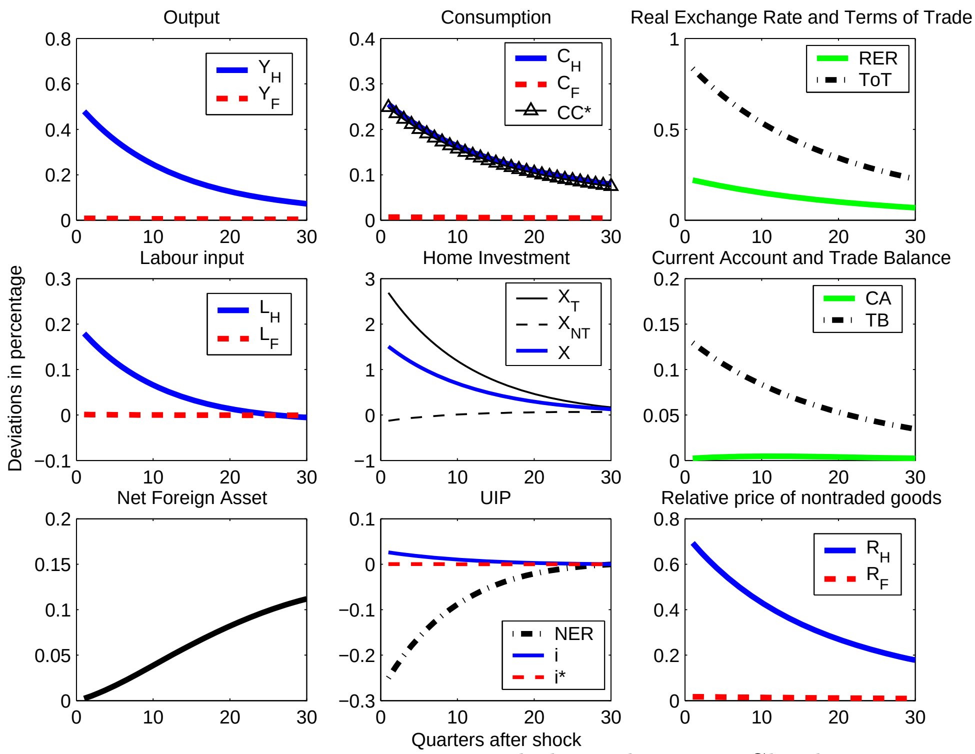


Figure 3: Response to Traded Productivity Shock

Note: The impulse responses show percentage deviation from steady state from period 1 when there is a 1% shock to traded productivity: RER - real exchange rate; CA, TB - current account and trade balance measured as percentage of output; NER - nominal exchange rate; i, i* - interest rate of small open economy and US; CC* - relative consumption to US; subscript H denotes home country whereas F denotes foreign country; subscript T denotes traded sector and NT denotes non-traded sector.

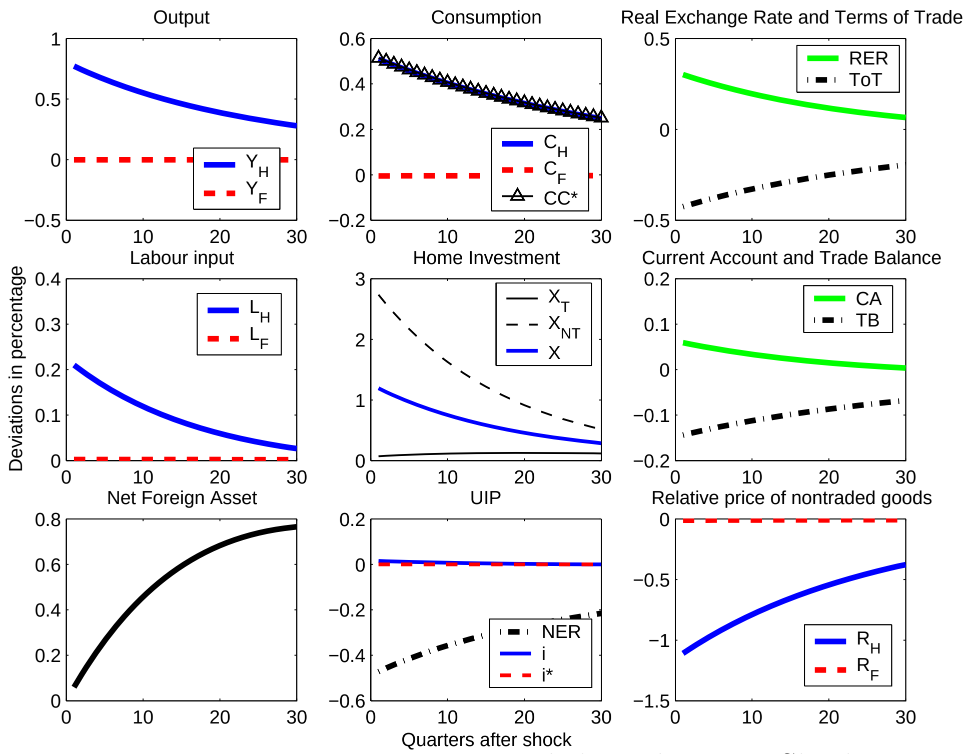


Figure 4: Response to Non-Traded Productivity Shock

Note: The impulse responses show percentage deviation from steady state from period 1 when there is a 1% shock to non-traded productivity: RER - real exchange rate; CA, TB - current account and trade balance measured as percentage of output; NER - nominal exchange rate; i, i* - interest rate of small open economy and US; CC* - relative consumption to US; subscript H denotes home country whereas F denotes foreign country; subscript T denotes traded sector and NT denotes non-traded sector.

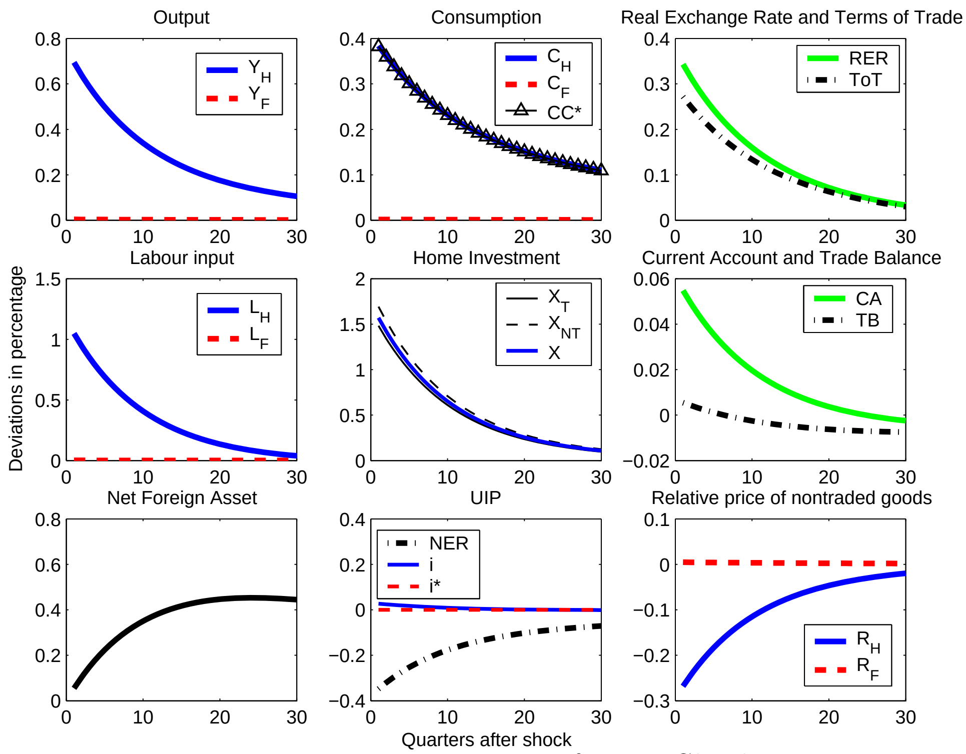


Figure 5: Response to Preference Shock

Note: The impulse responses show percentage deviation from steady state from period 1 when there is a 1% shock to preference: RER - real exchange rate; CA, TB - current account and trade balance measured as percentage of output; NER - nominal exchange rate; i, i* - interest rate of small open economy and US; CC* - relative consumption to US; subscript H denotes home country whereas F denotes foreign country; subscript T denotes traded sector and NT denotes non-traded sector.

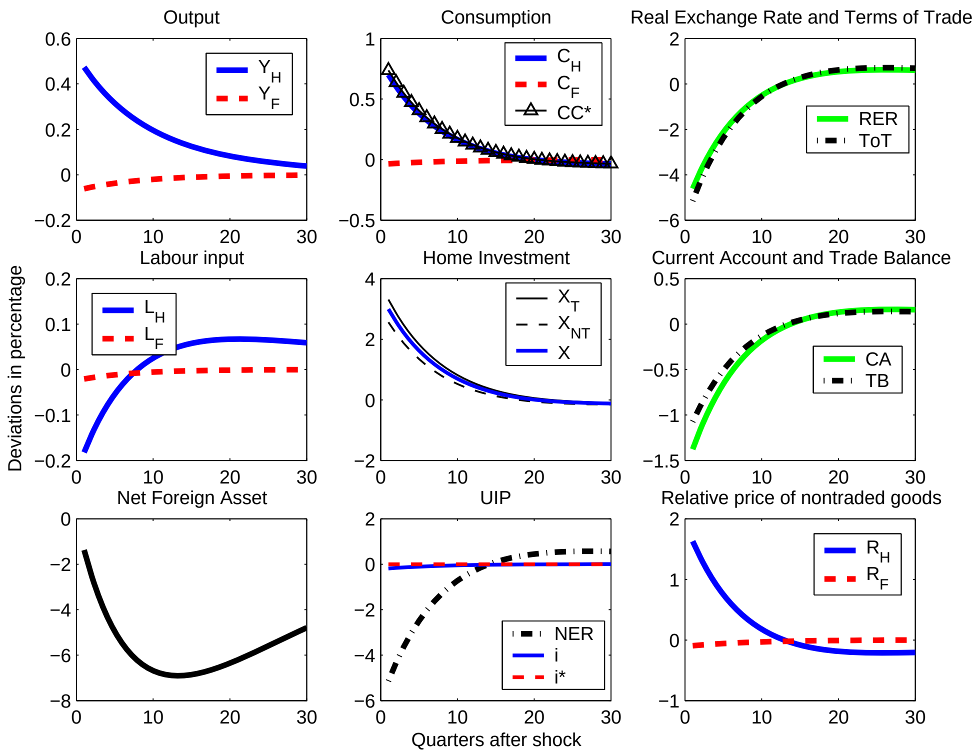


Figure 6: Response to UIP Deviation

Note: The impulse responses show percentage deviation from steady state from period 1 when there is a 1% shock to UIP: RER - real exchange rate; CA, TB - current account and trade balance measured as percentage of output; NER - nominal exchange rate; i, i* - interest rate of small open economy and US; CC* - relative consumption to US; subscript H denotes home country whereas F denotes foreign country; subscript T denotes traded sector and NT denotes non-traded sector.

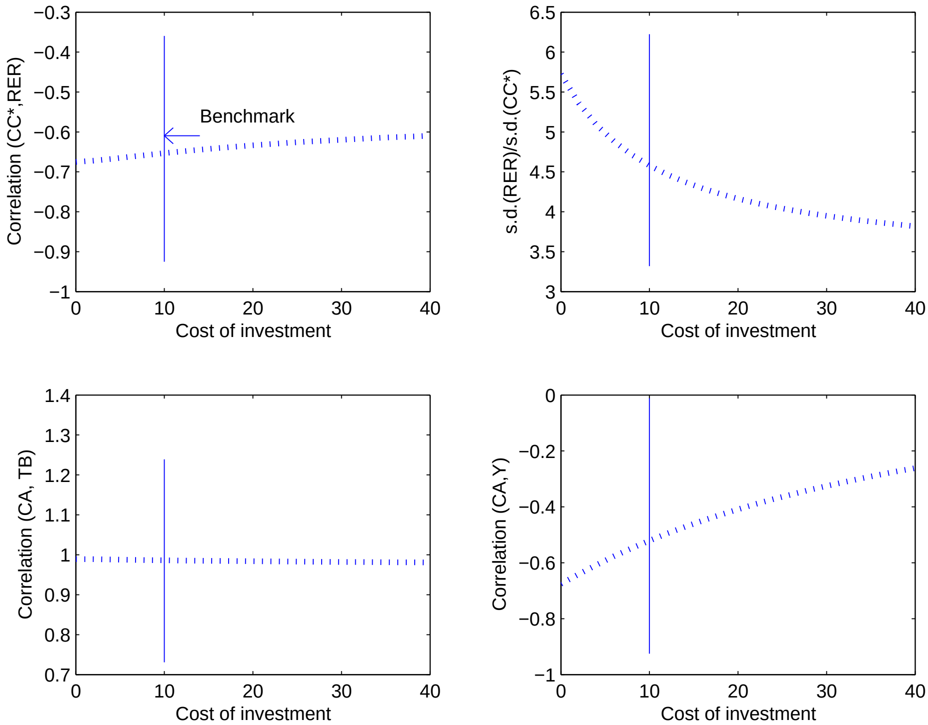


Figure 7: Sensitivity - Investment Cost (b)

Note: The charts show the sensitivity to the investment cost coefficient. The vertical line denotes the benchmark calibration of Table 1. The four charts show the key correlation and relative volatility statistics from calibrated model. CC* denotes relative consumption to US; RER: real exchange rate; CA: current account; TB: trade balance; Y: real output. CA and TB are measured as ratio to GDP.

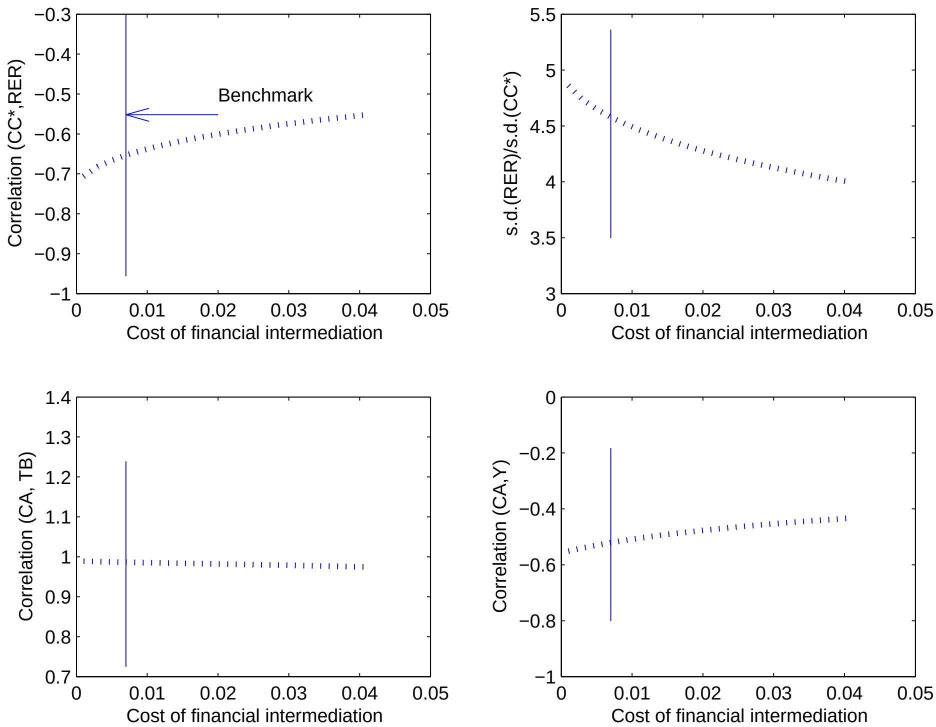


Figure 8: Sensitivity - Financial Intermediation Costs (ϵ)

Note: The charts show the sensitivity to the bond holding cost coefficient. The vertical line denotes the benchmark calibration of Table 1. The four charts show the key correlation and relative volatility statistics from calibrated model. CC* denotes relative consumption to US; RER: real exchange rate; CA: current account; TB: trade balance; Y: real output. CA and TB are measured as ratio to GDP.

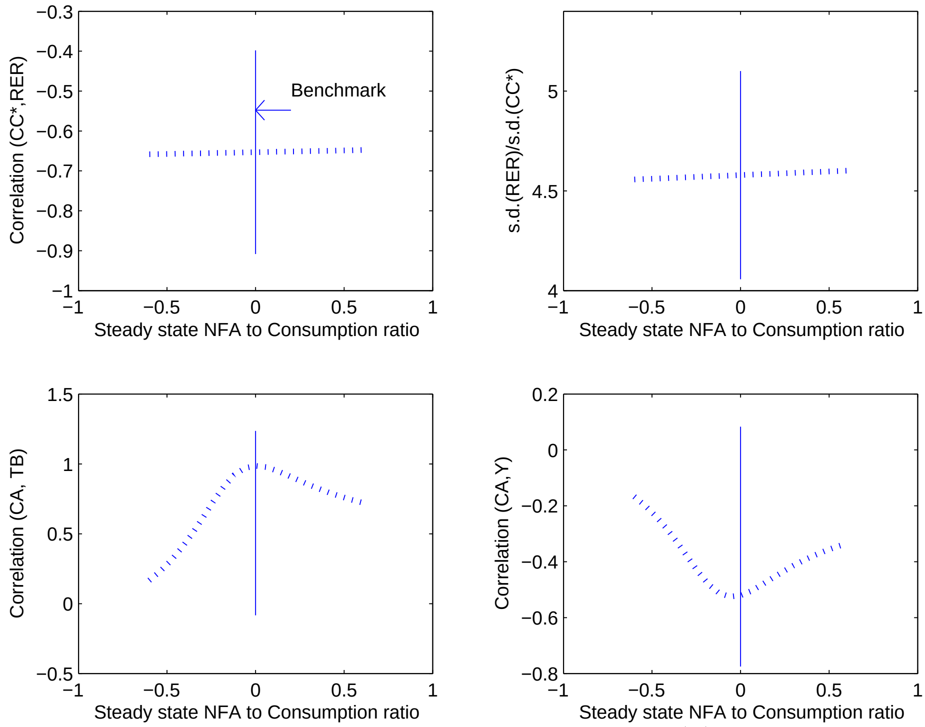


Figure 9: Sensitivity - NFA ratio ($\overline{a}$)

Note: The charts show the sensitivity to the steady state of net foreign asset position coefficient. The vertical line denotes the benchmark calibration of Table 1. The four charts show the key correlation and relative volatility statistics from calibrated model. CC* denotes relative consumption to US; RER: real exchange rate; CA: current account; TB: trade balance; Y: real output. CA and TB are measured as ratio to GDP.

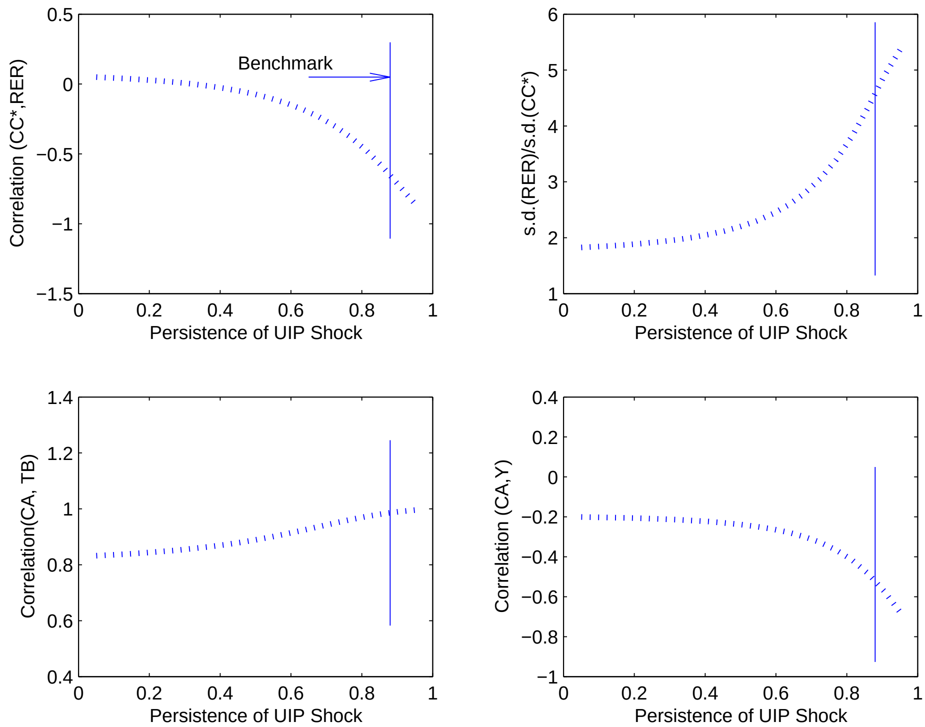


Figure 10: Sensitivity - UIP Deviation (ρ_{UIP})

Note: The charts show the sensitivity to the UIP shock persistence coefficient. The vertical line denotes the benchmark calibration of Table 1. The four charts show the key correlation and relative volatility statistics from calibrated model. CC* denotes relative consumption to US; RER: real exchange rate; CA: current account; TB: trade balance; Y: real output. CA and TB are measured as ratio to GDP.

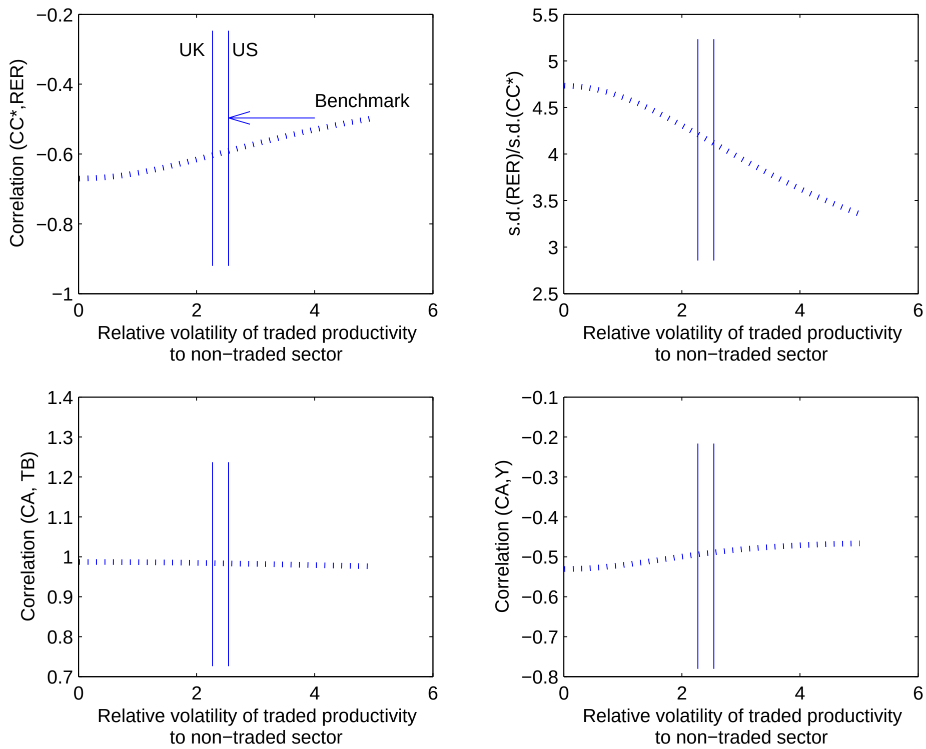


Figure 11: Sensitivity - Traded Sector Volatility (σ_A/σ_{AN})

Note: The charts show the sensitivity to the relative volatility coefficient. The vertical line denotes the benchmark calibration of Table 1. The four charts show the key correlation and relative volatility statistics from calibrated model. CC* denotes relative consumption to US; RER: real exchange rate; CA: current account; TB: trade balance; Y: real output. CA and TB are measured as ratio to GDP.

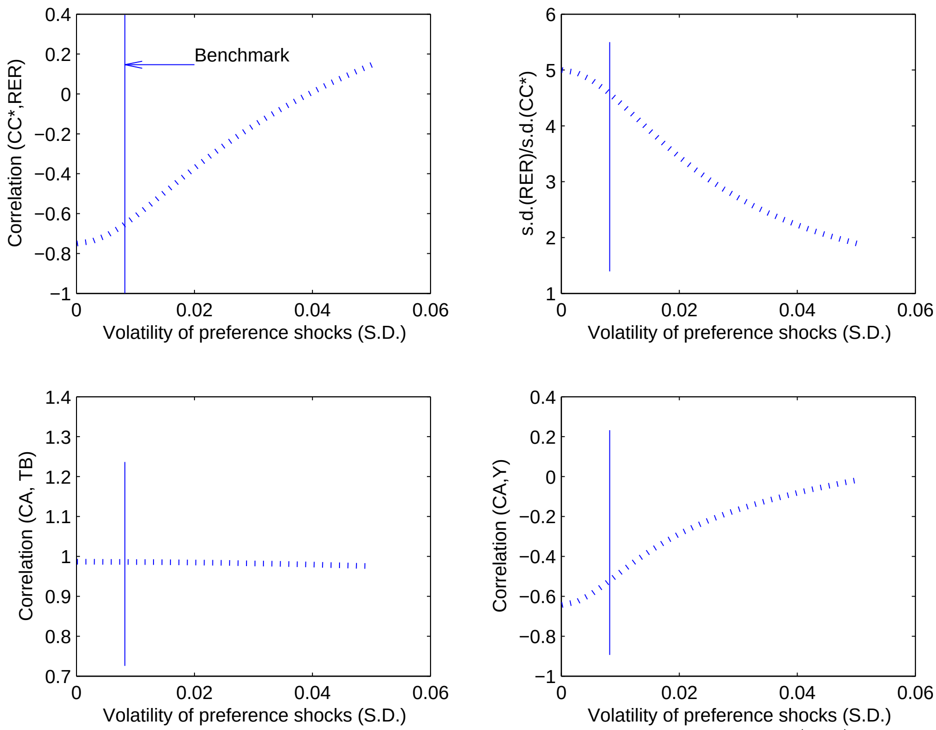


Figure 12: Sensitivity - Preference Shock Volatility (σ_P)

Note: The charts show the sensitivity to the volatility of preference shocks. The vertical line denotes the benchmark calibration of Table 1. The four charts show the key correlation and relative volatility statistics from calibrated model. CC* denotes relative consumption to US; RER: real exchange rate; CA: current account; TB: trade balance; Y: real output. CA and TB are measured as ratio to GDP.