A Tale of Two Indexes: Predicting Equity Market Downturns in China

Sébastien Lleo
William T. Ziemba

SRC Discussion Paper No 82
September 2018

Systemic Risk Centre

Discussion Paper Series

ISSN 2054-538X

Abstract
Predicting stock market crashes is a focus of interest for both researchers and practitioners. Several prediction models have been developed, mostly for use on mature financial markets. In this paper, we investigate whether traditional crash predictors, the price-to earnings ratio, the Cyclically Adjusted Price-to-Earnings ratio and the Bond-Stock Earnings Yield Differential model, predict crashes for the Shanghai Stock Exchange Composite Index and the Shenzhen Stock Exchange Composite Index. We also constructed active investment strategies based on these predictors. We found that these crash predictors have predictive power and the active strategies delivered lower risk and higher risk-adjusted return than a simple buy and hold investment.

Keywords: stock market crashes, Shanghai Stock Exchange, Shenzhen stock exchange, Bond-Stock Earnings Yield Differential (BSEYD), price earnings-ratio, Cyclically-Adjusted Price Earnings ratio (CAPE).

JEL Classification Numbers: G14, G15, G12, G10.

This paper is published as part of the Systemic Risk Centre's Discussion Paper Series. The support of the Economic and Social Research Council (ESRC) in funding the SRC is gratefully acknowledged [grant number ES/R009724/1].

Sébastien Lleo, Finance Department, NEOMA Business School
William T. Ziemba, University of British Columbia, Vancouver, BC and Systemic Risk Centre, London School of Economics

Published by
Systemic Risk Centre
The London School of Economics and Political Science
Houghton Street
London WC2A 2AE

A Tale of Two Indexes: Predicting Equity Market Downturns in China

Sébastien Lleo*and William T. Ziemba†

Abstract

Predicting stock market crashes is a focus of interest for both researchers and practitioners. Several prediction models have been developed, mostly for use on mature financial markets. In this paper, we investigate whether traditional crash predictors, the price-to-earnings ratio, the Cyclically Adjusted Price-to-Earnings ratio and the Bond-Stock Earnings Yield Differential model, predict crashes for the Shanghai Stock Exchange Composite Index and the Shenzhen Stock Exchange Composite Index. We also constructed active investment strategies based on these predictors. We found that these crash predictors have predictive power and the active strategies delivered lower risk and higher risk-adjusted return than a simple buy and hold investment.

JEL: G14, G15, G12, G10.

EconLit Subject Descriptors: G140, G150, G120, G100.

Keywords: stock market crashes, Shanghai Stock Exchange, Shenzhen stock exchange, Bond-Stock Earnings Yield Differential (BSEYD), price-earnings-ratio, Cyclically-Adjusted Price Earnings ratio (CAPE).

1 Introduction

The Chinese stock market is one of the most interesting equity markets in the world by its size, scope, structure and recency. These features have a

*Finance Department, NEOMA Business School, Email: sebastien.lleo@NEOMA-bs.fr

†University of British Columbia, Vancouver, BC and Systemic Risk Centre, London School of Economics

deep influence on its behavior and returns, including on the occurrence of rare events, in particular stock market crashes and downturns. In fact, the "2015 Chinese stock market crash" is just the latest in a series of 22 major downturns in a twenty-six years history.

The academic literature on bubbles and crashes is well established, starting with studies on bubbles by Blanchard and Watson (1982), Flood et al. (1986), Camerer (1989), Allen and Gorton (1993), Diba and Grossman (1988), Abreu and Brunnermeier (2003) and more recently Corgnet et al. (2015), Andrade et al. (2016) or Sato (2016). A rich literature on predictive models has also emerged. We can classify bubble and crash prediction models in three broad categories, based on the type of methodology and variable used: fundamental models, stochastic models and sentiment-based models.

Fundamental models use fundamental variables such as stock prices, corporate earnings, interest rates, inflation or GNP to forecast crashes. The Bond-Stock Earnings Differential (BSEYD) measure (Ziemba and Schwartz, 1991; Lleo and Ziemba, 2012, 2015, 2017) is the oldest model in this category, which also includes the CAPE (Lleo and Ziemba, 2017) and the ratio of the market value of all publicly traded stocks to the current level of the GNP (MV/GNP) that Warren Buffett popularized (Buffett and Loomis, 1999, 2001; Lleo and Ziemba, 2018).

Stochastic models construct a probabilistic representation of the asset prices, either as a discrete or continuous time stochastic process. Examples include the local martingale model proposed by Jarrow and Protter (Jarrow et al., 2011a; Jarrow, 2012; Jarrow et al., 2011b,c), the disorder detection model proposed by Shiryaev, Zhitlukhin and Ziemba (Shiryaev and Zhitlukhin, 2012a,b; Shiryaev et al., 2014, 2015) and the earthquake model of Gresnigt et al. (2015). When it comes to actual implementation, the local martingale model and the disorder detection model share the same starting point: they assume that the evolution of the asset price $S(t)$ can be best described using a geometric Brownian motion:

$$dS(t) = \mu(t, S(t))S(t)dt + \sigma(t, S(t))S(t)dW(t), S(0) = s_0, t \in \mathbb{R}^+$$

where $W(t)$ is a standard Brownian motion on the underlying probability space. However, the two models look at different aspects of the evolution. The disorder detection model detects crashes by looking for a change in regime in the drift μ and volatility σ. The local martingale model detects bubbles by testing whether the volatility σ is a local martingale or a strict martingale. In contrast, the earthquake model uses a jump-diffusion process and has a shorter forecasting horizon: 5 trading days. Gresnigt et al. (2015) implement the Epidemic-type Aftershock Sequence model (ETAS) geophysics model proposed by Ogata (1988), based on an Hawkes process, a type of inhomogeneous point process.

Behavioral models look at crashes in relation to market sentiment and behavioral biases. Goetzmann et al. (2016) use surveys of individual and institutional investors, conducted regularly over a 26 year period in the United States, to assess the subjective probability of a market crash and investigate the effect of behavioral biases on the formulation of these subjective probabilities. This research takes its roots in recent efforts to measure investor sentiment on financial markets (Fisher and Statman, 2000, 2003; Baker and Wurgler, 2006) and identify collective biases such as overconfidence and excessive optimism (Barone-Adesi et al., 2013).

Bubble and crash prediction models have a well documented track record of anecdotal successes on particular events, but until recently no systematic statistical methodology existed to test empirically their predictive ability (Lleo and Ziemba, 2017). In this paper, we test statistically the ability of the three main fundamental models - the BSEYD, P/E ratio and CAPE - to predict stock market downturns on Chinese markets. We leave aside Warren Buffett's ratio of the market value of all publicly traded stocks to the current level of the GNP (MV/GNP) simply because this measure cannot be computed frequently enough to provide a meaningful sample of data in the context of Chinese markets.

We choose to focus on the Chinese stock market, and more precisely on the two leading indexes - the Shanghai Stock Exchange Composite Index (SHCOMP) and the Shenzhen Stock Exchange Composite Index (SZECOMP) - because the Chinese stock market is arguably one of the most interesting equity markets in the world by its sheer size, scope, structure and recency. These features have a deep influence on its behavior and returns, including the occurrence of rare events such as stock market crashes and downturns. These characteristics make Chinese markets a particularly challenging and insightful testing environment for fundamental models that have, by and large, been designed using data from established Western equity markets. In the process, we gain new insights into both the statistical behavior of the Chinese market and the relevance of crash prediction models.

2 A Brief Overview of the Chinese Stock Market

Mainland China has two main stock exchanges, the Shanghai Stock Exchange (SSE) and the Shenzhen Stock Exchange (SZSE). The Shanghai Stock Exchange is the larger of the two. With an average market capitalization of USD 3.715 billion over the first half of 2016, it is the fourth largest stock market in the world[1]. The modern Shanghai Stock Exchange came into existence on November 26, 1990 and started trading on December 19, 1990. The Shenzhen Stock Exchange was founded on December 1, 1990, and started trading on July 3, 1991. While the largest and most established companies usually trade on the Shanghai Stock Exchange, the Shenzhen Stock Exchange is home to smaller and privately-owned companies.

With an average market capitalization of USD 6.656 billion over the first half of 2016, the Shanghai and Shenzhen Stock Exchanges taken together represent the third largest stock market in the world after the New York Stock Exchange at USD 17.970 billion, and the NASDAQ at USD 6.923 billion, and before 4th place Japan Exchange Group at USD 4.625 billion and fifth place LSE Group at USD 3.598 billion[2].

On November 17, 2014, the Chinese government launched the Shanghai-Hong Kong Stock Connect to enable investors in either market to trade shares on the other market. The Hong Kong Exchanges and Clearing is currently the 8th largest exchange in the world with an average market capitalization of USD 2.932 Billion over the first half of 2016[3]. This announcement was followed by the creation of a Shenzhen-Hong Kong link on August 16th, 2016. These initiatives herald a closer integration between securities markets in China and further boosts the rapid development of the Chinese market.

Chinese companies may list their shares under various schemes, either domestically or abroad. Domestically, companies may issue:

- *A-shares*: common stocks denominated in Chinese Reminbi and listed on the Shanghai or Shenzhen stock exchanges.

- *B-shares*: special purpose shares denominated in foreign currencies but

[1] Source: The World federation of Exchanges, `http://www.world-exchanges.org/home/index.php/statistics/monthly-reports` retrieved on September 13th, 2016

[2] Source: *ibid*

[3] Source: *ibid*

listed on the domestic stock exchange. Until 2001, only foreign investors had access to B-shares.

In addition to B-shares, foreign investors interested in the Chinese equity market may buy:

- *H-shares*: shares denominated in Hong Kong Dollars and traded on the Hong Kong Stock Exchange.
- *L-chips, N-chips and S-chips*: shares of companies with significant operations in China, but incorporated respectively in London, New York and Singapore.
- *American Depository Receipts (ADRs)*: an ADR is a negotiable certificate issued by a U.S. bank representing a specified number of shares in a foreign stock traded on an American exchange. As of October 2015, there were about 110 Chinese ADRs listed on American exchanges and another 200 Chinese ADRs on American over-the-counter markets.

The diversity of investment schemes available shows that although the Shanghai and Shenzhen Stock Exchange are a large and crucial part of the Chinese equity market, they do not represent the whole market. For example, there are also *red chips* (shares of companies incorporated outside mainland China but owned or substantially controlled by Chinese state-owned companies) and *P-chips* (shares of companies owned by private individuals and traded outside mainland China, for example on the Hong Kong stock exchange).

Our study focuses on equity market downturns on the Shanghai and Shenzhen Stock Exchanges.

3 Six Main Stylized Facts

The unique history and structure of the Chinese stock markets has a direct effect on the behaviour and statistical properties of the two leading equity indexes: the SHCOMP and SZECOMP. The SHCOMP and SZECOMP are market capitalization weighted indexes of shares listed respectively on the SSE and SZSE, respectively. In August 2016, the SHCOMP SZECOMP consisted of the shares of 1,155 and 478 Chinese companies.

We observe and discuss six main stylized facts on the historical distribution of daily log returns on the SHCOMP and SZECOMP. Collectively, these stylized facts indicate that the SHCOMP and SZECOMP behave differently from the mature equity markets in Europe and North America.

3.1 Stylized Fact 1: The return distribution is highly volatile, right skewed with very fat tails

The daily log return on the SHCOMP from December 20, 2017 until June 30, 2016 averaged 0.0541%, with a median return of 0.0693%. The lowest and highest daily returns were respectively -17.91% and +71.92%. Table 1 also gives the corresponding statistics at a weekly and monthly frequency. The returns are highly volatile: the standard deviation of daily returns is 2.40%, equivalent to around 40 times the mean daily return. The distribution of daily returns is positively skewed (skewness = 5.26) with surprisingly fat tails (Kurtosis = 149). As a result, the Jarque-Bera statistic is 5,419,808, rejecting normality at any level of significance. The Jarque-Bera statistic also leads to a strong rejection of normality for weekly and monthly data. The aggregational gaussianity, the tendency for the empirical distribution of log-returns to get closer to normality as the time scale Δt over which the returns are calculated increases, is much weaker on the SHCOMP and SZECOMP than on the S&P500 where Cont (2001) initially documented it.

We make similar observations on the SZECOMP. Table 1 shows that over the entire period, the daily log return on the SZE averaged 0.04784%, with a median return of 0.05933%. The lowest and highest daily returns were respectively -23.36% and +27.11%. Here as well, the returns are highly volatile: the standard deviation of daily returns is 2.28%, equivalent to around 50 times the mean daily return. The distribution of daily returns has a mildly positive skewness (skewness = 0.3517) and very fat tails (Kurtosis = 17). The Jarque-Bera statistic for the SZECOMP still reaches 52,879. The test leads to a rejection of normality at any level of significance not only for daily data, but also for weekly and monthly data.

3.2 Stylized Fact 2: The SHCOMP and SZECOMP do not exhibit a strong dependence structure

We turn our attention to the joint behavior of the SHCOMP and SZECOMP during the period from April 4, 1991 to June 30, 2016 (6,170 daily observa-

Descriptive Statistics	SHCOMP			SZECOMP		
	Daily	Weekly	Monthly	Daily	Weekly	Monthly
Number of observations	6,242	1,318	308	6,235	1,291	302
Mean	0.0541%	0.2497%	1.0326%	0.04784%	0.2345%	1.0644%
Median	0.0693%	0.0652%	0.7122%	0.05933%	0.1938%	0.8864%
Minimum	-17.9051%	-22.6293%	-37.3283%	-23.3607%	-33.5690%	-31.2383%
Maximum	71.9152%	90.0825%	101.9664%	27.2210%	51.9035%	60.9060%
Standard deviation	2.3848%	5.5872%	12.8898%	2.2808%	5.1795%	11.5411%
Variance	0.000569	0.000031	0.000166	0.000520	0.002683	0.013320
Skewness	5.1837	5.3543	2.3414	0.3517	1.2229	0.8724
Kurtosis	148.5003	78.5864	20.7742	17.2496	17.2522	6.6661
Jarque-Bera statistics	5,534,005	320,053	4,336	52,879.47	11,248.32	207.43
(p-value)	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$	$(< 2.2e-16)$

Table 1: **Descriptive statistics for daily, weekly and monthly log returns on the SHCOMP and SZECOMP**

tions). We compute the Pearson linear correlation, Spearman's rho (rank correlation) and Kendal's tau of the daily log returns. While the Pearson linear correlation measures the strength of the linear dependence of two data series, Spearman's rho computes the correlation between data of the same rank, and Kendal's Tau measures the distance between two ranking lists based on pairwise disagreements. Spearman's rho and Kendall's tau are non parametric: they do not require any assumption on the underlying distribution. At 0.6801, 0.7922 and 0.6443 respectively, the Pearson linear correlation, Spearman's rho and Kendall's Tau are all statistically different from 0. However, neither of them is close to 1. In fact, the statistical association between the SHCOMP and the SZECOMP is noticeably weaker than, for example, the association between the S&P500 and the NASDAQ. Over the same period, the two US indices had respective Pearson linear correlation, Spearman's rho and Kendall's Tau of 0.8742, 0.8592 and 0.6884.

3.3 Stylized Fact 3: The tail behavior of the SHCOMP and SZECOMP can be modeled using a Generalized Pareto Distribution

Extreme Value Theory (EVT) is the method of choice to uncover the statistical properties of rare events. We analyze the tail behavior of the SHCOMP and SZECOMP. We refer the reader to Coles (2001) for a concise and clear introduction to EVT and to Embrechts et al. (2011) for a thorough tour of the subject.

Here, we apply EVT to the loss distribution, which we define as the negative of the probability distribution of returns, meaning that if a stock index returns -1.5% on a given day, the associated loss will be 1.5%. We focus on the tail behavior, identified as the loss above a given threshold u, that we will determine during our analysis. Let X be the random variable representing the loss, and let F be its cumulative density function. Then the cumulative density function of the loss in excess of u is:

$$F_u(y) = P(X - u \le y | X > u) = \frac{F(u+y) - F(u)}{1 - F(u)},$$

for $0 \le y \le x_F - u$, where x_F is the right endpoint of F.

Theorem 3.1 (Pickands-Balkema-de Haan (PBH) (Pickands, 1975; Balkema and de Haan, 1974))**.** *For a large class of distribution functions F, and for u large enough, we can approximate the conditional excess distribution $F_u(y)$ by a Generalized Pareto Distribution (GPD) $G_{\xi,\sigma}$, that is:*

$$F_u(y) \approx G_{\xi,\sigma}(y), \text{ where } G_{\xi,\sigma}(x) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma} y\right)^{-\frac{1}{\xi}} & \text{if } \xi \neq 0, \\ 1 - e^{-\frac{\xi}{\sigma}} & \text{if } \xi = 0, \end{cases}$$

for y in $[0, x_F - u]$ if $\xi \ge 0$ and $y \in \left[0, -\frac{\sigma}{\xi}\right]$ if $\xi < 0$.

The parameters σ and ξ are respectively the scale and shape parameter of the GPD.

There is no firm rule governing the choice of threshold u. This choice of threshold must achieve a trade-off. If u is to low then the PBH theorem will not apply. If u is too high, then we will have too few observations to estimate the parameters of the GPD accurately. For example, we have 6,242 daily return observations for the SHCOMP, out of which 2,851 correspond to negative returns (i.e. positive loss). We still have 716 observations at a threshold of 2%, and 128 at a threshold of 5% but only 48 at 7%. The situation is similar on the SZECOMP. A popular method to determine u consists in plotting the sample mean excess loss against the threshold u, and picking the threshold u such that the sample mean excess loss is broadly linear for $v \ge u$. Figure 1 displays the excess loss against threshold for both the SHCOMP and SZECOMP. For the SHCOMP, we observe that the sample mean excess loss against the threshold becomes broadly linear in the threshold u starting at about $u = 4\%$. At that level, we still have 211 observations to fit the Generalized Pareto distribution. For the SZECOMP, the

	SHCOMP	SZECOMP
Threshold	4	6
Number of observations	211	85
Scale parameter (standard error)	1.8214 (0.1821)	1.7141 (0.2829)
Shape parameter (standard error)	0.1292 (0.0731)	0.2176 (0.1266)
AIC	734	303
BIC	740	307

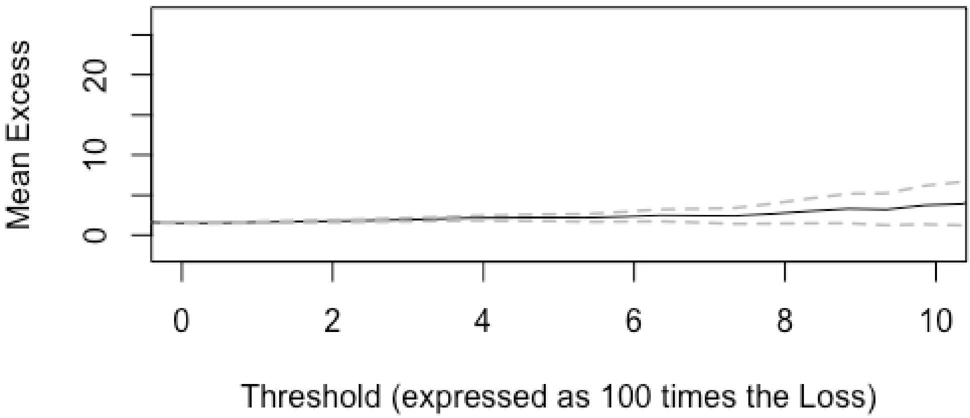
Table 2: Parameters of the Generalized Pareto distribution fitted to the tail of the SHCOMP and SZECOMP. The estimation is performed via maximum likelihood against 100× the loss to improve numerical stability.

post suggests choosing $u = 6\%$, which leaves us with 85 observations to fit the distribution.

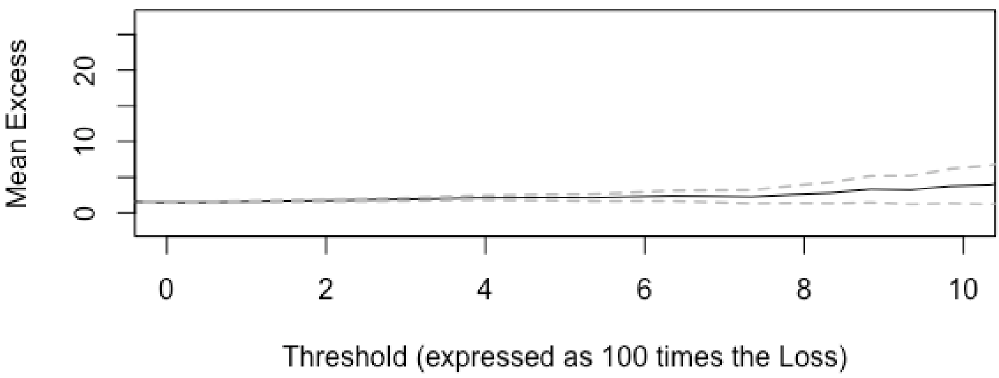
Finally, we estimate the scale parameter σ shape parameter ξ of the GPD using maximum likelihood. This estimation is performed against $100y$, or 100 times the loss, in order to improve numerical stability. Table 2 presents the estimated parameters, standard error of estimates as well as the Akaike Information Criterion (AIC) and the Bayesian Information Criterion (BIC) for both indexes.

3.4 Stylized Fact 4: Log returns do not exhibit a significant autocorrelation

Figures 2 show that the autocorrelation of daily log returns up to lag 20 are in the interval [- 0.03, 0.06]. This suggests that neither indexes exhibits a short-term memory: today's return does not help forecast tomorrow' s return. An analysis of the PACF leads to similar conclusions.

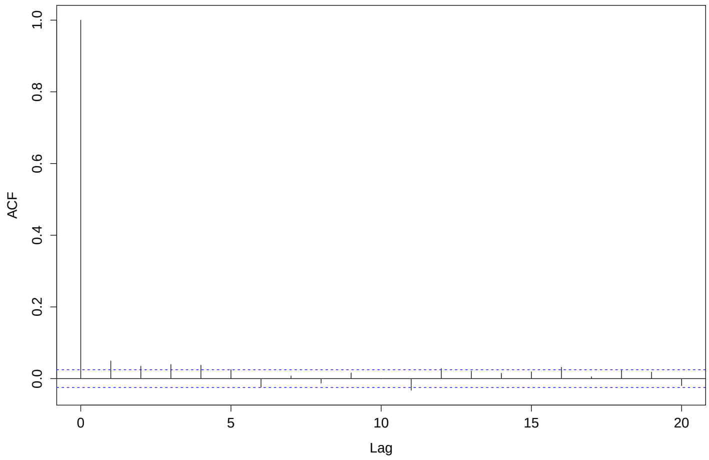


(a)

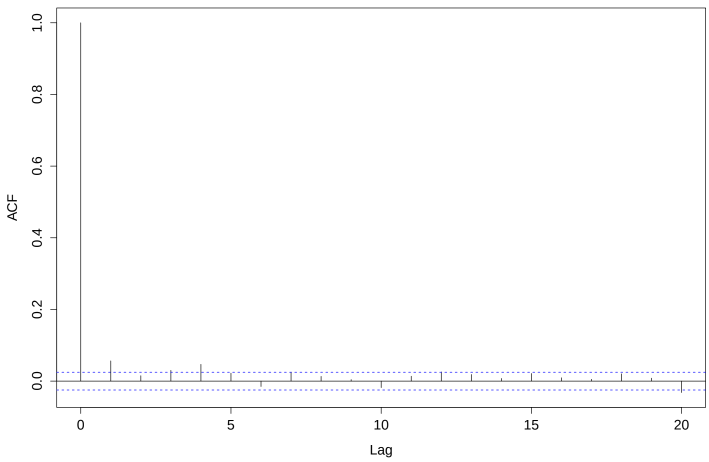


(b)

Figure 1: Sample mean excess loss against the threshold for the SHCOMP (a) and SZECOMP (b).



(a) SHCOMP



(b) SZECOMP

Figure 2: Sample autocorrelation of the daily log returns on the SHCOMP and SZECOMP up to lag 20

3.5 Stylized Fact 5: A Gaussian Hidden Markov Chain provides a good probabilistic description of the evolution of log returns... but we need between five and six states.

Stylized Fact 1 indicates that the distribution of log returns is skewed with fat tails, while Stylized Fact 2 supports the use of a Markov model to describe the probabilistic behavior of the log returns on the SHCOMP and SZECOMP. We look for a simple discrete-time Markov Model able to describe the probabilistic behavior and the evolution of log returns.

A good starting point is to look at Hidden Markov Models (HMMs). HMMs are a useful way to model the behavior of a physical or economic system when we suspect that this behavior is determined by the transition between a finite number of unobservable "regimes" or "states." We refer the reader to the excellent presentation of HMMs in Rabiner (1989) and Rabiner and Juang (1993).

The simplest, and often the best, HMM models are Gaussian Hidden Markov Chains. In these models, the returns in each state are conditionally normally distributed. The parameters of each normal distribution are specific to that state. As the state transitions over time, the returns are drawn from different normal distributions, resulting in an aggregate distribution that bears little resemblance to a normal distribution. Gaussian HMMs are estimated via the Baum-Welch algorithm (Baum et al., 1970), an application of the well-known EM algorithm (see Dempster et al., 1977).

One of the difficulties is to find the optimal number of states for the model. To that end, it is customary to use an information criterion such as the AIC or the BIC to discriminate between model formulations. The optimal model minimize the absolute value of the information criterion. Contrary to the LogLikelihood, the AIC and BIC penalize the model for the number of parameters used. This penalty is stiffer in the BIC than in the AIC.

Tables 3 present the Loglikelihood, AIC and BIC for HMMs with one to seven states, fitted respectively on the SHCOMP and the SZECOMP. We performed the numerical procedure using the *depmixS4* package in R. For the SHCOMP, we find that the optimal model specification, the specification that minimizes the AIC and BIC, is a six-state model, while the optimal model for the SZECOMP is a slightly more parsimonious, but still large, five-state model. By contrast, a two or three-state model usually proves ad-

	1	2	3	4	5	6	7
				SHCOMP			
LogLikelihood	14,464	16,514	16,827	16,887	16,895	17,183	17,194
AIC	-28,924	-33,013	-33,625	-33,728	-33,723	-34,273	-34,265
BIC	-28,910	-32,966	-33,531	-33,573	-33,494	-33,956	-33,847
Number of parameters	2	7	14	23	34	47	62
				SHCOMP			
LogLikelihood	14,726	16,053	16,225	16,298	16,331	16,346	16,389
AIC	-29,447	-32,091	-32,422	-32,550	-32,593	-32,598	-32,653
BIC	-29,434	-32,044	-32,328	-32,395	-32,364	-32,598	-32,235
Number of parameters	2	7	14	23	34	47	62

Table 3: Hidden Markov Model fitting for the daily log returns on the SHCOMP and SZECOMP

State	Initial Probability	Mean	Standard Deviation
1	0.00	-1.0996%	1.4419%
2	0.00	1.6517%	9.5992%
3	0.00	0.3433%	1.0724%
4	0.00	0.2015%	2.0513%
5	1.00	-0.1354%	0.7037%
6	0.00	0.1464%	3.6882%

Table 4: Initial probability and parameters of the Gaussian distributions for each state of the HMM

equate for mature indexes such as the S&P 500.

The transition probability matrix P_{SHCOMP} for the SHCOMP is

$$\begin{pmatrix} 7.2689e-01 & 4.3972e-175 & 2.6749e-231 & 2.7311e-01 & 6.2691e-303 & 2.3976e-220 \\ 2.5311e-04 & 9.3881e-01 & 2.2433e-58 & 3.8282e-02 & 2.0163e-05 & 2.2637e-02 \\ 2.5622e-202 & 1.2809e-90 & 8.6320e-01 & 1.4495e-117 & 5.2279e-33 & 1.3680e-01 \\ 8.5411e-03 & 1.1560e-01 & 9.9862e-04 & 8.7043e-01 & 5.1730e-27 & 4.4282e-03 \\ 3.6201e-127 & 2.5515e-02 & 4.1153e-22 & 1.6460e-15 & 5.2982e-01 & 4.4466e-01 \\ 2.3777e-115 & 6.5345e-06 & 1.1122e-02 & 2.8003e-47 & 7.5704e-01 & 2.3184e-01 \end{pmatrix}.$$

The initial probability and the parameters of the normal distribution for each state are given in Table 4.

State	Initial Probability	Mean	Standard Deviation
1	0.00	-1.2627%	1.5456%
2	1.00	-0.0750%	0.7600%
3	0.00	0.1433%	7.1826%
4	0.00	0.3734%	1.2108%
5	0.00	0.1057%	2.7757%

Table 5: Initial probability and parameters of the Gaussian distributions for each state of the HMM

The transition probability matrix $P_{SZECOMP}$ for the SZECOMP is

$$\begin{pmatrix} 2.9883e-01 & 2.4262e-16 & 1.1294e-27 & 6.7930e-01 & 2.1866e-02 \\ 6.0465e-02 & 9.3154e-01 & 7.9941e-03 & 1.1369e-18 & 5.6279e-14 \\ 4.3118e-16 & 1.7914e-02 & 7.8243e-01 & 3.9601e-62 & 1.9965e-01 \\ 1.4320e-01 & 2.6999e-02 & 3.1915e-03 & 8.1727e-01 & 9.3390e-03 \\ 1.3235e-08 & 2.9150e-03 & 1.8946e-02 & 2.5725e-02 & 9.5241e-01 \end{pmatrix}.$$

The initial probability and the parameters of the normal distribution for each state are given in Table 5.

3.6 Stylized Fact 6: Downturns and large market movements occur frequently

The return distribution of the SHCOMP has fat tails, which indicates that extreme events are more likely to occur than a Normal distribution would predict. Here, we focus on the large downward movements that occurred on the SHCOMP and SZECOMP.

Earlier studies, such as Lleo and Ziemba (2015, 2017), defined an equity market downturn or crash as a decline of at least 10% from peak to trough based on the closing prices for the day, over a period of at most one year (252 trading days). We identify a correction on the day when the daily closing price crosses the 10% threshold. The identification algorithm is as follows:

1. *Identify all the local troughs in the data set.* Today is a local trough if there is no lower closing price within ±30 business days.

2. *Identify the crashes.* Today is a crash identification day if all of the following conditions hold:

(a) The closing level of the index today is down at least 10% from its highest level within the past year, and the loss was less than 10% yesterday;

(b) This highest level reached by the index prior to the present crash differs from the highest level corresponding to a previous crash;

(c) This highest level occurred after the local trough that followed the last crash.

The objective of these rules is to guarantee that the downturns we identify are distinct. Two downturns are not distinct if they occur within the same larger market decline. Although these rules might be argued with, they have the advantage of being unambiguous, robust and easy to apply.

A total of 22 downturns occurred on the SHCOMP between December 19, 1990 and June 30, 2016. These downturns are presented in Table 6. On average, the downturns lasted 163 days and had a 27.8% decline in the value of the index. With 22 downturns in 25 years, the SHCOMP had as many downturns as the S&P 500 over the 50 year period from January 31, 1964 to December 31, 2014.

A total of 21 downturns occurred on the SZECOMP between April 3, 1991 and June 30, 2016. They are presented in Table 7. On average, the downturns lasted 122 days and had a 26.4% decline in the value of the index. While the number and magnitude of equity market corrections are comparable between both indexes, we observe that downturns tend to last noticeably longer on average on the Shanghai stock Exchange than on the Shenzhen Stock Exchange.

Collectively, these stylized facts indicate that the SHCOMP and SZECOMP behave differently from the mature equity markets in Europe and North America.

4 Methodology

4.1 Signal Construction

The construction process for the signal and hit sequence is crucial to ensure that the crash prediction models produce out of sample predictions free from look-ahead bias. It also eliminates data snooping by setting the parameters *ex ante*, with no possibilities of changing them when we construct the hit

	Crash Identification Date	Peak Date	SHCOMP Index at Peak	Trough date	SHCOMP Level at trough	Peak-to-trough decline (%)	Peak-to-trough duration (in days)
1	1992-05-27	1992-05-25	1421.57	1992-11-17	393.52	72.3%	176
2	1993-02-23	1993-02-15	1536.82	1993-03-31	925.91	39.8%	44
3	1994-09-19	1994-09-13	1033.47	1995-02-7	532.49	48.5%	147
4	1996-08-26	1996-07-24	887.6	1996-09-12	757.09	14.7%	50
5	1996-11-6	1996-10-28	1022.86	1996-12-24	865.58	15.4%	57
6	1997-05-16	1997-05-12	1500.4	1997-09-23	1041.97	30.6%	134
7	1998-08-7	1998-06-3	1420	1998-08-17	1070.41	24.6%	75
8	1999-07-1	1999-06-29	1739.21	1999-12-27	1345.35	22.6%	181
9	2000-09-22	2000-08-21	2108.69	2000-09-25	1875.91	11%	35
10	2001-02-21	2001-01-10	2125.62	2001-02-22	1907.26	10.3%	43
11	2001-07-30	2001-06-13	2242.42	2002-01-22	1358.69	39.4%	223
12	2003-04-23	2002-07-8	1732.93	2003-11-18	1316.56	24%	498
13	2004-04-29	2004-04-6	1777.52	2004-09-13	1260.32	29.1%	160
14	2006-08-4	2006-07-11	1745.81	2006-08-7	1547.44	11.4%	27
15	2007-02-2	2007-01-24	2975.13	2007-02-5	2612.54	12.2%	12
16	2007-06-4	2007-05-29	4334.92	2007-07-5	3615.87	16.6%	37
17	2007-11-8	2007-10-16	6092.06	2008-11-4	1706.7	72%	385
18	2009-08-12	2009-08-4	3471.44	2009-08-31	2667.75	23.2%	27
19	2010-10-27	2009-11-23	3338.66	2011-01-25	2677.43	19.8%	428
20	2012-12-27	2012-03-2	2460.69	2013-06-27	1950.01	20.8%	482
21	2014-06-25	2013-09-12	2255.6	2014-06-25	2025.5	10.2%	286
22	2015-06-19	2015-06-12	5166.35	2015-08-26	2927.29	43.3%	75

Table 6: The SHCOMP Index experienced 22 crashes between December 19, 1990 and June 30, 2016.

	Crash Identification Date	Peak Date	SZECOMP Index at Peak	Trough date	SZECOMP Level at trough	Peak-to-trough decline (%)	Peak-to-trough duration (in days)
1	1992-06-3	1992-05-26	312.21	1992-06-16	233.73	25.1%	21
2	1993-03-5	1993-02-22	359.44	1993-07-21	203.91	43.3%	149
3	1996-05-10	1995-05-22	169.66	1996-08-26	152.55	10.1%	462
4	1996-09-10	1996-09-4	274.56	1996-12-24	242.01	11.9%	111
5	1997-05-16	1997-05-12	517.91	1997-09-23	312.73	39.6%	134
6	1998-07-6	1998-06-3	441.04	1998-08-18	317.1	28.1%	76
7	1999-07-1	1999-06-29	525.14	1999-12-27	395.69	24.7%	181
8	2000-09-25	2000-08-21	643.77	2000-09-25	578.76	10.1%	35
9	2001-02-8	2000-11-23	654.37	2001-02-22	568.26	13.2%	91
10	2001-07-30	2001-06-13	664.85	2002-01-22	371.79	44.1%	223
11	2004-04-26	2004-04-7	470.55	2004-09-13	315.17	33%	159
12	2006-08-2	2006-07-12	446.61	2006-08-7	380.26	14.9%	26
13	2007-06-1	2007-05-29	1292.44	2007-07-5	1015.85	21.4%	37
14	2007-10-25	2007-10-9	1551.19	2007-11-28	1219.98	21.4%	50
15	2008-01-22	2008-01-15	1576.5	2008-11-4	456.97	71%	294
16	2009-08-14	2009-08-4	1149.27	2009-09-1	900.53	21.6%	28
17	2009-12-22	2009-12-3	1234.17	2010-07-5	921.34	25.3%	214
18	2010-11-17	2010-11-10	1389.54	2011-01-25	1136.58	18.2%	76
19	2013-06-24	2013-05-30	1043.47	2013-06-25	879.93	15.7%	26
20	2014-03-28	2014-02-17	1160.39	2014-04-28	1007.27	13.2%	70
21	2015-06-19	2015-06-12	3140.66	2015-09-15	1580.26	49.7%	95

Table 7: The SZECOMP Index experienced 21 crashes between March 25, 1992 and June 30, 2016.

sequence. More importantly, the construction of the hit sequence removes the effect of autocorrelation, making it possible to test the accuracy of the measures using a standard likelihood ratio test.

Equity market crash prediction models such as the BSEYD, the high P/E model or the CAPE generate a signal to indicate that an equity market downturn is likely at a given horizon h. This signal occurs whenever the value of a crash measure crosses a threshold. Given a prediction measure $M(t)$, a crash signal occurs whenever

$$SIGNAL(t) = M(t) - K(t) > 0 \tag{4.1}$$

where $K(t)$ is a time-varying threshold for the signal.

Three parameters define the signal: (i) the choice of measure $M(t)$; (ii) the definition of threshold $K(t)$; and (iii) the specification of a time interval H between the occurrence of the signal and that of an equity market downturn.

We construct the measures using two time-varying thresholds: (i) a dynamic confidence interval based on a Normal distribution; and (ii) a dynamic confidence interval using Cantelli' s inequality - see Problem 7.11.9 in Grimmett and Stirzaker (2001) for a statement of the mathematical result, and Lleo and Ziemba (2012, 2017) for applications to crash predictions.

To construct the confidence intervals, we compute the sample mean and standard deviation of the distribution of the measures as a moving average and a rolling horizon standard deviation respectively. Using rolling horizon means and standard deviations has the advantage of providing data consistency. Importantly, this construction only makes use of information known at the time of the calculation. The h-day moving average at time t, denoted by μ_t^h, and the corresponding rolling horizon standard deviation σ_t^h are

$$\mu_t^h = \frac{1}{h}\sum_{i=0}^{h-1} x_{t-i}, \qquad \sigma_t^h = \sqrt{\frac{1}{h-1}\sum_{i=0}^{h-1}(x_{t-i} - \mu_t^h)^2}.$$

We establish the one-tailed confidence interval at the 95% level. This corresponds to 1.645 standard deviations above the mean in the Normal distribution.

We select the one-tailed confidence interval at $\alpha = 95\%$, corresponding to 1.645 standard deviations above the mean in the Normal distribution. This choice is consistent with the crash prediction literature and can be traced to the first published work on the BSEYD Ziemba and Schwartz (1991).

The historical development of statistical inference by Fisher, E. Pearson and Neyman, among others, has contributed to popularizing the choice of $\alpha = 95\%$ for two-tailed tests: R.A. Fisher suggested the use of a two-tailed 5% significance level (see for example pp. 45, 98, 104, 117 in Fisher, 1933; Neyman and Pearson, 1933; Neyman, 1934, 1937).

As an alternative to the normal confidence level, we construct the confidence level using Cantelli's inequality. This inequality relates the probability that the distance between a random variable X and its mean μ exceeds a number $k > 0$ of standard deviations σ to provide a robust confidence interval:

$$P\left[X - \mu \geq k\sigma\right] \leq \frac{1}{1+k^2}.$$

Setting $\beta := \frac{1}{1+k^2}$ yields $P\left[X - \mu \geq \sigma\sqrt{\frac{1}{\beta} - 1}\right] \leq \beta$. Contrary to the normal confidence level, Cantelli's inequality does not require any assumption on the shape of the underlying distribution. It should therefore provide more robust results for fat tailed distributions. The parameter β provides an upper bound for a one-tailed confidence level on any distribution. In our analysis, the horizon for the rolling statistics is h = 252 days. There is no clear rule on how to select β, so we chose $\beta = 25\%$ to produce a slightly higher threshold than the standard confidence interval. In a Normal distribution, we expect 5% of the observations to lie in the right tail, whereas Cantelli's inequality implies that the percentage of outliers in a distribution will be no higher than 25%.

The last parameter we need to specify is the horizon H. Earlier, we defined the crash identification time is the date by which the SHCOMP has declined by at least 10% in the last year (252 trading days). We define the local market peak as the highest level reached by the market index within 252 trading days before the crash. We set the horizon H to a maximum of 252 trading days prior to the crash identification date.

4.2 Signal Indicator and Crash Indicator

Crash prediction models have two components: (1) a signal indicator, which takes the value 1 or 0 depending on whether the measure has crossed the threshold, and (2) a crash indicator, which takes the value 1 when an equity market correction occurs and 0 otherwise. From a probabilistic perspective, these components are Bernoulli random variables, but they exhibit a high degree of autocorrelation, that is, a value of 1 (0) for the crash signal is more likely to be followed by another value of 1 (0) on the next day. This autocorrelation makes it difficult to test the accuracy of the model.

To remove the effect of autocorrelation, we define a signal indicator sequence $S = \{S_t, t = 1, \ldots, T\}$. This sequence records as the signal date the first day in a series of positive signals, and it only counts distinct signal dates. Two signals are distinct if a new signal occurs more than 30 days after the previous signal. The objective is to have enough time between two series of signals to identify them as distinct. The signal indicator S_t takes the value 1 if date t is the starting date of a distinct signal, and 0 otherwise. Thus, the event "a distinct signal starts on day t" is represented as $\{S_t = 1\}$. We express the signal indicator sequence as the vector $s = (S_1, \ldots, S_t, \ldots, S_T)$.

For the crash indicator, we denote by $C_{t,H}$ the indicator function returning 1 if the crash identification date of at least one equity market correction occurs between time t and time $t + H$, and zero otherwise. We identify the vector C_H with the sequence $C_H := \{C_{t,H}, t = 1, \ldots, T - H\}$ and define the vector $c_H := (C_{1,H}, \ldots, C_{t,H}, \ldots C_{T-H,H})$.

The number of correct predictions n is defined as

$$n = \#\{C_{t,H} = 1 | S_t = 1\} = \sum_{t=1}^{T} \mathbf{1}_{\{C_{t,H}=1|S_t=1\}},$$

where $\mathbf{1}_A$ is the indicator function returning 1 if condition A is satisfied, and 0 other wise.The accuracy of the crash prediction model is therefore the conditional probability $P(C_{t,H} = 1|S_t = 1)$ of a crash being identified between time t and time $t + H$, given that we observed a signal at time t. The higher the probability, the more accurate the model.

4.3 Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$ and Likelihood Ratio Test

We use maximum likelihood to estimate the conditional probability $P(C_{t,H} = 1|S_t = 1)$ and to test whether it is significantly higher than a random guess. We obtain a simple analytical solution because the conditional random variable $\{C_{t,H} = 1|S_t = 1\}$ is a Bernoulli trial with probability $p = P(C_{t,H} = 1|S_t = 1)$.

To estimate the probability p, we change the indexing to consider only events along the sequence $\{S_t|S_t = 1, t = 1, \dots T\}$ and denote by $X := \{X_i, i = 1, \dots, N\}$ the "hit sequence" where $x_i = 1$ if the ith signal is followed by a crash and 0 otherwise. Here N denotes the total number of signals, that is

$$N = \sum_{t=1}^{T} S_t$$

The sequence X can be expressed in vector notation as $x = (X_1, X_2, \dots, X_N)$. The empirical probability p is the ratio n/N.

The likelihood function L associated with the observations sequence X is

$$L(p|X) := \prod_{i=1}^{N} p^{X_i}(1-p)^{1-X_i}$$

and the log likelihood function $\mathcal{L}$ is

$$\mathcal{L}(p|X) := \ln L(p|X) = \sum_{i=1}^{N} X_i \ln p + \left(N - \sum_{i=1}^{N} X_i\right) \ln(1-p)$$

This function is maximized for $\hat{p} := \frac{\sum_{i=1}^{N} X_i}{N} = n/N$, so the maximum likelihood estimate of the probability $p = P(C_{t,H}|S_t)$ is the sample proportion of correct predictions.

We apply a likelihood ratio test to test the null hypothesis $H_0 : p = p_0$ against the alternative hypothesis $H_A : p \neq p_0$. The null hypothesis reflects the idea that the probability of a random, uninformed signal correctly predicting crashes is p_0. The probability p_0 is the probability to identify an equity market downturn within 252 days of a randomly selected period. To compute p_0 empirically, we tally the number of days that are at most 252 days before a crash identification date and divide by the total number of days in the sample.

A significant departure above p_0 indicates that the measure we are considering contains some information about future equity market corrections. The likelihood ratio is:

$$\Lambda = \frac{L(p = p_0|X)}{\max_{p\in(0,1)} L(p|X)} = \frac{L(p = p_0|X)}{L(p = \hat{p}|X)}. \tag{4.2}$$

The test statistic $Y := -2\ln\Lambda$ is asymptotically χ^2-distributed with $\nu = 1$ degree of freedom. We reject the null hypothesis $H_0 : p = p_0$ and accept that the model has some predictive power if $Y > c$, where c is the critical value chosen for the test. We perform the test for the three critical values 2.71, 3.84, and 6.63 corresponding respectively to a 90%, 95% and 99% confidence level.

4.4 Monte Carlo Study for Small Sample Bias

A limitation of this likelihood ratio test is that the χ^2 distribution is only valid asymptotically. In our case, the number of correct predictions follows a binomial distribution with an estimated probability of success $\hat{p}$ and N trials. However, "only" 18 downturns occurred during the period considered in this study: the continuous χ^2 distribution might not provide an adequate approximation for this discrete distribution. This difficulty is an example of small sample bias. We use Monte Carlo methods, with $K = 10,000$ paths, to obtain the empirical distribution of test statistics and address this bias.

4.5 Optimal Parameter Choice and Parameter Robustness

At a first glance, the statistical validity of the model seems to depend crucially on the signal construction, and therefore on two parameters: the confidence level α and the forecasting horizon H. The confidence level affects directly the number of signals that the model generates, and indirectly the accuracy of the model. The forecasting horizon influences the number of correct signals, as well as the uninformed probability p_0 used in the significance test, but it does not change the number of signals generated. It is easier to produce an accurate forecast if we have a longer horizon to prove us right than a shorter one. In this section, we set up a plan to test the sensitivity or robustness of the model to these two parameters.

First, we compute the optimal value for the confidence level α. We hold the time horizon H constant at 252 days, and seek the confidence level(s) $\alpha \in [0.9, 1]$ that maximizes the empirical accuracy $\hat{p}$:

$$\mathcal{A} = \text{argmax}_{\alpha \in [0.9,1)} \hat{p}(\alpha; H)$$

We are actually interested in the lowest confidence interval for which we $\hat{p} = 100\%$, as well as in the general evolution of the number of predications as the confidence level increases. We expect the accuracy of the measure to increase with the confidence level.

We are also interested in whether the model remains significantly better than a random guess if we chose a confidence level at the lower end of the confidence range. Answering this question will give us an indication on the robustness of the model in relation to a change or misspecification in the confidence level. This approach is an application of the robust likelihood statistics proposed by Lleo and Ziemba (2017) to a case where we test the robustness with respect to a single parameter.

Next, we look for the optimal value for the forecasting horizon H. We hold the confidence level α constant at 95% and look for the time horizon(s) H that maximizes the empirical accuracy $\hat{p}$:

$$\mathcal{H} = \text{argmax}_{H \in \{63,126,189,252\}} \hat{p}(H; \alpha)$$

We limit the range of our analysis to up to 252 days after the signal. Note the we cannot test the robustness of the model with respect to a change in forecasting horizon using the robust likelihood statistics proposed by Lleo and Ziemba (2017) because changing the forecasting horizon will affect the uninformed probability p_0.

4.6 Further Analysis: Measure-On-Measure Significance Test

Lleo and Ziemba (2017) proposed a likelihood ratio test for pairwise measure-on-measure comparisons to determine whether the accuracy p_i of a given model i is significantly higher than the estimated accuracy $\hat{p}_j = \text{argmax}_{p_j \in (0,1)} L(p_j|X)$ of model j.

We performed this test for all the measures studied in this paper, but did not find any significant difference between the accuracy of the measures at a 95% confidence level. This result indicates that none of the measures is statistically more accurate than the others.

5 The Price-to-Earnings Ratio

5.1 Scope of the Study

Practitioners have used the price-to-earnings (P/E) ratio to gauge the relative valuation of stocks and stock markets since at least the 1930s (for example, Graham and Dodd, 1934, discuss the use of the P/E ratio in securities analysis and valuation).

In this section, we test the predictive ability of the P/E ratio calculated using current earnings. The advantage of this definition for the SHCOMP is that it is available over the entire period from December 19, 1990 to June 30, 2016, a total of 6,243 daily observations. The same is not true for the SZECOMP. earnings and therefore P/E are only available starting July 2, 2001, a total of 3,640 daily observations.

5.2 Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$ and Likelihood Ratio Test

Table 8 shows that the P/E and logarithm of the P/E generated a total of 18 signals (based on normally distributed confidence intervals) and 19 signals (based on Cantelli' s inequality) on the SHCOMP. The number of correct predictions across models reaches 16 to 17. The accuracy of the models is in the narrow range from 88.89% to 89.47%. The type of confidence interval - normal distribution or Cantelli' s inequality - only have a minor influence on the end result.

Next, we test the accuracy of the prediction on the SHCOMP statistically. To apply the likelihood ratio test, we need to compute the uninformed prior probability p_0 that a day picked at random will precede a crash identification date by 252 days or less. We find that this probability is very high, at $p_0 = 69.57\%$. This finding is consistent with the stylized facts discussed in Section 2. The Likelihood ratio test indicates that both the P/E ratio and the logarithm of the P/E ratio are significant predictors of equity market downturns markets at the 90% confidence level. Moreover, the P/E ratio, computed using a standard confidence interval, and the log P/E ratio, based on Cantelli's inequality, are significant at the 95% confidence level. Thus, we cannot rule out that the P/E and log P/E/ have helped predict equity market downturns over the period.

The P/E and logarithm of the P/E generated a total of 8 to 9 signals, with 7 to 8 correct signals on the the SZECOMP. The accuracy of the models is in a narrow range from 87.50% to 88.89%. Here as well, the type of confidence interval - normal distribution or Cantelli' s inequality - only have a minor influence on the end result.

5.3 Monte Carlo Study for Small Sample Bias

We continue our analysis with a Monte Carlo test for small sample bias, presented in Table 9. We compute the critical values at the 90%, 95% and 99% confidence level for the empirical distribution. Because we only have a limited number of signals, the distribution is lumpy, making it difficult to obtain meaningful p-values. Still, we find that the Monte Carlo analysis is in broad agreement with our earlier conclusions about significance of the P/E ratio and its logarithm, as both measures are significant at the 90% confidence level. We conclude that small sample bias only has a very small effect on these measures and on their statistical significance.

The uninformed prior probability p_0 that a day picked at random will precede a crash identified date by 252 days or less is 58.49%. The Likelihood ratio test indicates that both P/E ratio measures and the logarithm of the P/E ratio calculated using a standard confidence interval are significant predictors of equity market downturns markets at the 95% confidence. The remaining measure, the logarithm of the P/E ratio calculated with Cantelli's inequality is significant at the 90% confidence level. The results of the Monte Carlo analysis, presented in Table 9, indicate that small sample bias only has a minor effect on the statistical significance of the measures. All the measures are still significant at the 90% confidence level.

5.4 Optimal Parameter Choice and Parameter Robustness

We follow up with an analysis of the sensitivity of the measures to a change of confidence level α and forecasting horizon H.

Model	Total number of signals	Number of correct predictions	ML Estimate $\hat{p}$	$L(\hat{p})$	Likelihood ratio Λ	Test statistics $-2\ln\Lambda$	p-value
				SHCOMP			
PE (confidence)	19	17	89.47%	1.67E-03	0.1159	4.3100*	3.79%
PE (Cantelli)	18	16	88.89%	1.88E-03	0.1486	3.8131†	5.09%
logPE (confidence)	18	16	88.89%	1.88E-03	0.1486	3.8131†	5.09%
logPE (Cantelli)	19	17	89.47%	1.67E-03	0.1159	4.31*	3.79%
				SZECOMP			
PE (confidence)	9	8	88.89%	4.33E-02	0.1313	4.0607*	4.39%
PE (Cantelli)	9	8	88.89%	4.33E-02	0.1313	4.0607*	4.39%
logPE (confidence)	9	8	88.89%	4.33E-02	0.1313	4.0607*	4.39%
logPE (Cantelli)	8	7	87.5%	4.91E-02	0.1980	3.2387†	7.19%

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 8: **SHCOMP and SZECOMP: Maximum likelihood estimate and likelihood ratio test for the PE and logPE** The Total Number of Signals is calculated as the sum of all the entries of the indicator sequence S. The Number of Correct Predictions is the tally of crashes preceded by the signal. It is calculated as the sum of all the entries of the indicator sequence X. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. $L(\hat{p})$ is the likelihood of the crash prediction model, computed using the maximum likelihood estimate $\hat{p}$. The likelihood ratio $\Lambda = \frac{L(p_0|X)}{L(p=\hat{p}|X)}$ is the ratio of the likelihood under the null hypothesis $p = p_0$ to the likelihood using the estimated probability $\hat{p}$. The estimated test statistics, equal to $-2\ln\Lambda$, is asymptotically χ^2-distributed with 1 degree of freedom. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance and the p-value indicated in the table are both based on this distribution. The critical values at the 95%, 99% and 99.5% level are respectively 3.84, 6.63 and 7.88.

Model	Total number of signals	ML Estimate $\hat{p}$	Critical Value			Test statistics $-2\ln\Lambda(p_0)$
			90% confidence	95% confidence	99% confidence	
	SHCOMP					
PE (confidence)	19	89.47%	2.38	4.31	7.61	4.3100†
PE (Cantelli)	18	88.89%	2.38	4.31	7.61	3.8131†
logPE (confidence)	18	88.89%	2.99	3.81	6.99	3.8131†
logPE (Cantelli)	19	89.47%	2.99	3.81	6.99	4.3100*
	SZECOMP					
PE (confidence)	9	88.89%	2.31	4.06	4.92	4.0607†
PE (Cantelli)	9	88.89%	2.31	4.06	4.92	4.0607†
logPE (confidence)	9	88.89%	2.31	4.06	8.86	4.0607†
logPE (Cantelli)	8	87.50%	2.31	4.06	8.86	3.2387†

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 9: **SHCOMP: Monte Carlo likelihood ratio test for the PE and logPE** The Total Number of Signal is calculated as the sum of all the entries of the indicator sequence S. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. Colums 4 to 6 report the critical values at the 95%, 99% and 99.5% confidence level for the empirical distribution generated using $K = 10,000$ Monte-Carlo simulation. The test statistics in column 7 is equal to $-2\ln\Lambda(p_0) = -2\ln\frac{L(p_0|X)}{L(p=\hat{p}|X)}$ and that in column 9 is $-2\ln\Lambda\left(\frac{1}{2}\right) = -2\ln\frac{L(\frac{1}{2}|X)}{L(p=\hat{p}|X)}$. The level of significance indicated for both tests are based on the empirical distribution. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance indicated in the test statistics column and the p-value indicated in the table are both based on and empirical distribution generated through Monte-Carlo simulations.

5.4.1 Shanghai

We focus here on measures computed using a standard confidence interval for clarity as we would obtain similar results for measures computed using Cantelli's inequality. Table 10 reports the key statistics of the measure for various confidence levels. Picking a confidence level at the low end of our range, $\alpha = 90\%$, the P/E ratio generates 22 signal while the log P/E produces 21 signals. With an accuracy of 81.82%, the P/E ratio is no longer significant at the 90% confidence level. On the other hand, the log P/E is 85.71% accurate, maintaining itself above the critical value corresponding to a 90% confidence level. Expanding the scope of our investigation outside of the initial [0.9,1) range to consider a broader confidence range of $[0.8, 1)$, we find that the accuracy and significance of the P/E ratio and log P/E ratio broadly increase with the confidence level, while the number of signals decreases monotonically, as expected. In fact, the accuracy of the models reaches 100% at $\alpha = 0.99$ for the P/E ratio and $\alpha = 0.97$ for the log of the P/E, but with only 14 to 15 predictions out of 22 crashes.

The increase in the accuracy and significance is not monotonic because of the limited number of predictions: adding one correct prediction or one incorrect predictions tends to have a noticeable impact on the accuracy of the measure. This makes the transitions lumpy rather than smooth. Still, we observe that both the P/E ratio and the log P/E ratio remain significant at the 90% confidence level in the range $[0.925, 1)$, suggesting that the two measures are not overly sensitive to a small change in the confidence parameter α.

We conclude our analysis of the P/E and log P/E by investigating the sensitivity of these measures to a change in horizon H. Table 11 reports the key statistics for $H = 63, 126, 189$ and 252 days, corresponding to 3 months, 6 months, 9 months and 1 year. The accuracy of the signals decreases as we we shorten the time horizon, and so does the uninformed probability p_0. Overall, the P/E and log P/E become significant when the horizon reaches 9 months to 1 year, and their test statistics reaches its maximum at 9 months.

5.4.2 Shenzhen

An analysis of the sensitivity of the measure to a change in the confidence parameter α produces a surprising outcome. Contrary to what we observed with the SHCOMP, the results for the SZECOMP, presented in Table 12, show that the accuracy of the measures, and therefore their statistical sig-

Confidence	0.8	0.85	0.9	0.925	0.95	0.975	0.99
				P/E ratio			
Number of signals	21	21	22	22	19	16	15
Number of correct signals	15	18	18	19	17	15	15
Proportion of correct signals	71.43%	85.71%	81.82%	86.36%	89.47%	93.75%	100%
Test statistics	0.0348	$2.9770^{\dagger}$	1.7190	$3.4022^{\dagger}$	4.3100^{*}	5.7847^{*}	-
p-value	85.2%	8.45%	18.98%	6.51%	3.79%	1.62%	-
				logP/E ratio			
Number of signals	21	21	21	19	18	14	11
Number of correct signals	15	17	18	17	16	14	11
Proportion of correct signals	71.43%	80.95%	85.71%	89.47%	88.89%	100%	100%
Test statistics	0.0348	1.4050	$2.9770^{\dagger}$	4.3100^{*}	$3.8131^{\dagger}$	-	-
p-value	85.2%	23.59%	8.45%	3.79%	5.09%	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 10: **SHCOMP: Accuracy and statistical significance of the P/E ratio and logP/E ratio as a function of the confidence level** α. The numbers presented in this table are based on a forecasting horizon $H = 252$ days. With this choice, the uninformed probability that a random guess would correctly identify an equity market downturn is $p_0 = 67.64\%$ Row 1,2 and 3 respectively report the total number of signals generated by the P/E ratio, the number of correct signals, and the proportion of correct signals computed as the ratio of the number of correct signals to the total number of signals. Rows 4 and 5 respectively report the test statistics and p-value for the P/E ratio. The subsequent rows present the same information for the log P/E ratio.

Horizon (days)	63	126	189	252
Uninformed probability p_0	50.99%	59.59%	66.41%	69.57%
		P/E ratio		
Number of correct signals	18	18	18	18
Proportion of correct signals	57.89%	73.68%	89.47%	89.47%
Test statistics	0.3648	1.6561	5.4937^*	4.31^*
p-value	54.58%	19.81%	1.91%	3.79%
		logP/E ratio		
Number of correct signals	19	19	19	19
Proportion of correct signals	66.67%	77.78%	88.89%	88.89%
Test statistics	1.8093	2.6753	4.904^*	$3.8131^\dagger$
p-value	17.86%	10.19%	2.68%	5.09%

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 11: **SHCOMP: Accuracy and statistical significance of the P/E ratio and log P/E ratio as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, both the P/E ratio generated 19 signals, and the log P/E ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.The subsequent rows present the same information for the log P/E ratio.

Confidence	0.8	0.85	0.9	0.925	0.95	0.975	0.99
				P/E ratio			
Number of signals	12	12	9	11	9	7	7
Number of correct signals	11	11	8	10	8	6	6
Proportion of correct signals	91.67%	91.67%	88.89%	90.91%	88.89%	85.71%	85.71%
Test statistics	6.6736**	6.6736**	4.0607*	5.7831*	4.0607*	2.4528	2.4528
p-value	0.98%	0.98%	4.39%	1.62%	4.39%	11.73%	11.73%
				logP/E ratio			
Number of signals	11	10	9	10	9	8	6
Number of correct signals	10	9	8	9	8	7	5
Proportion of correct signals	90.91%	90.00%	88.89%	90.00%	88.89%	87.50%	83.33%
Test statistics	5.7831**	4.9107*	4.0607*	4.9107*	4.0607*	3.2387†	1.7150
p-value	1.62%	2.67%	4.39%	2.67%	4.39%	7.19%	19.03%

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 12: **SZECOMP: Accuracy and statistical significance of the P/E ratio and logP/E ratio as a function of the confidence level** α. The numbers presented in this table are based on a forecasting horizon $H = 252$ days. With this choice, the uninformed probability that a random guess would correctly identify an equity market downturn is $p_0 = 58.49\%$ Row 1,2 and 3 respectively report the total number of signals generated by the P/E ratio, the number of correct signals, and the proportion of correct signals computed as the ratio of the number of correct signals to the total number of signals. Rows 4 and 5 respectively report the test statistics and p-value for the P/E ratio. The subsequent rows present the same information for the log P/E ratio.

nificance, declines overall as the α increases. The accuracy of the models decline from 91.67% at $\alpha = 80\%$ to 85.71% at $\alpha = 99\%$. This is enough to push the p-value up from 0.98% to 11.73%. This counterintuitive outcome is a result of the fact that the total number of signals generally decrease, as α increases. This is what we observe here: the models generate 11 to 12 signals at $\alpha = 80\%$ but only 6 to 7 at $\alpha = 99\%$. Since the models are already particularly accurate, an erroneous signal therefore results in a larger loss of accuracy at $\alpha = 99\%$ than at $\alpha = 80\%$.

The measures do not exhibit a high sensitivity to a change in the time horizon H. The results of the analysis, summarized in Table 13, show that the models remain significant at the 90% confidence level across all four time horizons: 63 days, 126 days, 189 days and 252 days.

6 The Cyclically-Adjusted Price-to-Earnings Ratio and the Bond-Stocks Earnings Yield Differential Model

6.1 Scope of the Study

The P/E ratio calculated using current earnings might be overly sensitive to current economic and market conditions. Graham and Dodd (1934) warned against this risk and advocated the use of a P/E ratio based on average earnings over ten years. In their landmark survey, Campbell and Shiller (1988) found that the R^2 of a regression of log returns on the S&P 500 with a 10 year horizon against the log of the price-earnings ratio computed using average earnings over the previous 10 and 30 years equals 0.566 and 0.401 respectively, hinting at a link between average past earning and future stock prices. This later led Shiller to suggest the use of a Cyclically Adjusted Price-to-Earnings ratio (CAPE), or a price-to-earnings ratio using 10-year average earnings, to forecast the evolution of the equity risk premium (Shiller, 2005).

The BSEYD, the second model we test, relates the yield on stocks (measured by the earnings yield, which is also the inverse of the P/E ratio) to that on nominal Government bonds.

$$BSEYD(t) = r(t) - \rho(t) = r(t) - \frac{E(t)}{P(t)}, \tag{6.1}$$

where $\rho(t)$ is the earnings yield at time t and $r(t)$ is the current 10-year government bond yield $r(t)$. The BSEYD was initially developed for the Japanese market in 1988, shortly before the stock market crash of 1990, based on the 1987 stock market in the US (Ziemba and Schwartz, 1991). The BSEYD has since been used successfully on a number of international markets (see the review article Lleo and Ziemba, 2015), and the 2007-2008 SHCOMP meltdown (Lleo and Ziemba, 2012).

We test the forecasting ability of four measures:

1. **PE0**: P/E ratio based on current earnings. This is the measure we tested in Section 5;

2. **CAPE10**: CAPE, which is a P/E ratio computed using average earnings over the previous 10-years;

Horizon (days)	63	126	189	252
Uninformed probability p_0	19.56%	35.03%	48.79%	58.49%
		P/E ratio		
Number of correct signals	9	9	9	9
Proportion of correct signals	44.44%	66.67%	77.78%	88.89%
Test statistics	2.8646†	3.7184†	3.189†	4.0607*
p-value	9.05%	5.38%	7.41%	4.39%
		logP/E ratio		
Number of correct signals	8	8	8	8
Proportion of correct signals	44.44%	66.67%	77.78%	88.89%
Test statistics	2.8646†	3.7184†	3.189†	4.0607*
p-value	9.05%	5.38%	7.41%	4.39%

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 13: **SZECOMP: Accuracy and statistical significance of the P/E ratio and log P/E ratio as a function of the forecasting horizon** H**.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, both the P/E ratio generated 19 signals, and the log P/E ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.The subsequent rows present the same information for the log P/E ratio.

3. **BSEYD0**: BSEYD based on current earnings;
4. **BSEYD10**: BSEYD using average earnings over the previous 10-years.

We also test the logarithm of these measures: **logPE0**, **logCAPE10**, **logBSEYD0** and **logBSEYD10**. The logBSEYD is defined as:

$$logBSEYD(t) = \ln \frac{r(t)}{\rho(t)} = \ln r(t) - \ln \frac{E(t)}{P(t)}. \tag{6.2}$$

Because the CAPE10 and BSEYD10 require 10 years of earnings data, and the Bloomberg data series for 10-year government bonds only starts on October 31, 2006, we cannot use the full range of stock market data. The analysis in this section covers the period between October 31, 2006 and June 30, 2016. Over this period, the SHCOMP experienced seven declines of more than 10%, while the SZECOMP had nine.

We omit from the discussion results related to Cantelli's inequality because of space constraints. These results are nearly identical to the results we obtain for measures based on a standard confidence interval.

6.2 Maximum Likelihood Estimate of $p = P(C_{t,H}|S_t)$ and Likelihood Ratio Test

Table 14 displays results for the eight measures calculated with a confidence interval based on a normal distribution on both stock market indexes.

Looking at the SHCOMP, none of the measures produced more than 5 signals. The CAPE, logCAPE and BSEYD10 generated 3 signals each. The accuracy of the measures reaches a low of 40% for logBSEYD0 and a high of 100% for CAPE10 and logCAPE10. Only five of the eight measures are 75% accurate or better. By comparison, the uninformed prior probability that a day picked at random will precede a crash identification date by 252 days or less is $p_0 = 70.99\%$. Because of the relatively short period and small number of downturns, only CAPE10 and logCAPE10 appear significant. However, these two models only predicted three of the six crashes.

Overall, none of the models perform convincingly on the SHCOMP. The PE0 and logPE0 ratio, which we found to be significant predictors over the entire dataset in the previous section, are not significant over this restricted time period. With a 75% accuracy, they have a small edge over the uniformed prior p_0, but this edged is not significant. What's more, the BSEYD-based models do not perform as well as the P/E-based models. This is a puzzle because the BSEYD model contains additional information that is not in the P/E, namely government bond yields.The BSEYD and logBSEYD models have also been shown to perform better than the P/E ratio and CAPE on the American market (Lleo and Ziemba, 2017).

The situation on the SZECOMP is markedly different: all the measures, but one, have a 100% accuracy on the six or seven signals that they generated. The remaining measure, logBSEYD10, had six correct predictions out of seven signals, which implies a 85.71% accuracy. Although this is much higher than the uniformed prior p_0 at about 67%, the sample is to small for the difference in accuracy to be statistically significant. The discrepancy between the results observed on the SHCOMP and SZECOMP raises a number of questions. Is the difference in accuracy merely statistical, resulting from the small number of equity market downturns in the sample, or does it reveal a divergence in the microstructure of the two indexes? While the results computed in Section 5 for the P/E ratio seem to hint at the former, the latter is also a possibility, especially in light of the second Stylized Fact in Section 3.2.

6.3 Monte Carlo Study for Small Sample Bias

The results of the Monte Carlo analysis for small sample bias, presented in table 15 support the conclusions of the asymptotic maximum likelihood test. In the case of the SZECOMP, the Monte Carlo analysis for small bias is not informative because most measure have an infinite test statistic.

6.4 Optimal Parameter Choice and Parameter Robustness

Finally, we explore the sensitivity of the measures to a change in the forecasting horizon H and confidence α.

Table 16 reports the results of an analysis of the measures' sensitivity to a change in the confidence parameter α on the Shanghai market. To the exception of the logBSEYD0, the accuracy of the measures increase as α increases. Five measures out of eight reach a 100% accuracy at $\alpha = 97.5\%$, over two or three signals. The logBSEYD0 and logBSEYD10 remain the

Signal Model	Total number of signals	Number of correct predictions	ML Estimate $\hat{p}$	$L(\hat{p})$	Likelihood ratio Λ	Test statistics $-2\ln\Lambda$	p-value
				SHCOMP			
BSEYD0	4	3	75.00%	1.05E-01	0.717	0.6654	41.47%
logBSEYD0	5	2	40.00%	3.46E-02	0.7901	0.4713	49.24%
PE0	4	3	75.00%	1.05E-01	0.717	0.6654	41.47%
logPE0	4	3	75.00%	1.05E-01	0.717	0.6654	41.47%
BSEYD10	3	2	66.67%	1.48E-01	0.9228	0.1606	68.86%
logBSEYD10	5	3	60.00%	3.46E-02	0.9778	0.0449	83.23%
CAPE10	3	3	100.00%	-	-	-	-
logCAPE10	3	3	100.00%	-	-	-	-
				SZECOMP			
BSEYD0	6	6	100.00%	-	-	-	-
logBSEYD0	7	7	100.00%	-	-	-	-
PE0	6	6	100.00%	-	-	-	-
logPE0	6	6	100.00%	-	-	-	-
BSEYD10	7	6	85.71%	5.67E-02	0.5266	1.2826	25.74%
logBSEYD10	7	7	100.00%	-	-	-	-
CAPE10	6	6	100.00%	-	-	-	-
logCAPE10	5	5	100.00%	-	-	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 14: **SHCOMP and SZECOMP: Maximum likelihood estimate and likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm**The Total Number of Signals is calculated as the sum of all the entries of the indicator sequence S. The Number of Correct Predictions is the tally of crashes preceded by the signal. It is calculated as the sum of all the entries of the indicator sequence X. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. $L(\hat{p})$ is the likelihood of the crash prediction model, computed using the maximum likelihood estimate $\hat{p}$. The likelihood ratio $\Lambda = \frac{L(p_0|X)}{L(p=\hat{p}|X)}$ is the ratio of the likelihood under the null hypothesis $p = p_0$ to the likelihood using the estimated probability $\hat{p}$. The estimated test statistics, equal to $-2\ln\Lambda$, is asymptotically χ^2-distributed with 1 degree of freedom. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance and the p-value indicated in the table are both based on this distribution. The critical values at the 95%, 99% and 99.5% level are respectively 3.84, 6.63 and 7.88.

Signal Model	Total number of signals	ML Estimate $\hat{p}$	Critical Value			Test statistics $-2\ln\Lambda(p_0)$
			90% confidence	95% confidence	99% confidence	
	SHCOMP					
BSEYD0	4	75%	4.74	4.74	6.44	0.6654
logBSEYD0	5	40%	2.62	5.92	8.05	0.4713
PE0	4	75%	4.74	4.74	6.44	0.6654
logPE0	4	75%	4.74	4.74	6.44	0.6654
BSEYD10	3	66.67%	3.55	4.83	4.83	0.1606
logBSEYD10	5	60%	2.62	5.92	8.05	0.0449
CAPE10	3	100.00%	3.55	4.83	4.83	-
logCAPE10	3	100%	3.55	4.83	4.83	-
	SZECOMP					
BSEYD0	6	100.00%	4.81	4.81	6.48	-
logBSEYD0	5	100.00%	4.31	5.61	5.61	-
PE0	6	100.00%	4.81	4.81	6.48	-
logPE0	6	100.00%	4.81	4.81	6.48	-
BSEYD10	7	85.71%	4.31	5.61	5.61	1.2826
logBSEYD10	7	100.00%	4.31	5.61	5.61	-
CAPE10	6	100.00%	4.81	4.81	6.48	-
logCAPE10	5	100.00%	4.01	4.01	4.66	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 15: **SHCOMP and SZECOMP: Monte Carlo likelihood ratio test for the BSEYD0, PE0, BSEYD10 and CAPE10 and their logarithm** The Total Number of Signal is calculated as the sum of all the entries of the indicator sequence S. The Maximum Likelihood estimate $\hat{p}$ is the probability of correctly predicting a crash that maximises the likelihood function of the model. It is equal to the ratio of the number of correct prediction to the total number of signals. Colums 4 to 6 report the critical values at the 95%, 99% and 99.5% confidence level for the empirical distribution generated using $K = 10,000$ Monte-Carlo simulation. The test statistics in column 7 is equal to $-2\ln\Lambda(p_0) = -2\ln\frac{L(p_0|X)}{L(p=\hat{p}|X)}$ and that in column 9 is $-2\ln\Lambda\left(\frac{1}{2}\right) = -2\ln\frac{L(\frac{1}{2}|X)}{L(p=\hat{p}|X)}$. The level of significance indicated for both tests are based on the empirical distribution. The p-value is the probability of obtaining a test statistic higher than the one actually observed, assuming that the null hypothesis is true. The degree of significance indicated in the test statistics column and the p-value indicated in the table are both based on and empirical distribution generated through Monte-Carlo simulations.

	0.8	0.85	0.9	0.925	0.95	0.975	0.99	0.8	0.85	0.9	0.925	0.95	0.975	0.99
	BSEYD0							logBSEYD0						
Number of signals	7	6	6	4	4	2	2	6	5	5	5	5	4	4
Number of correct signals	4	3	3	3	3	2	2	3	2	2	2	2	2	2
Proportion of correct signals	57.14%	50%	50%	75%	75%	100%	100%	50%	40%	40%	40%	40%	50%	50%
Test statistics	0.0095	0.0681	0.0681	0.6654	0.6654	-	-	0.0681	0.4713	0.4713	0.4713	0.4713	0.0454	0.0454
p-value	92.22%	79.42%	79.42%	41.47%	41.47%	-	-	79.42%	49.24%	49.24%	49.24%	49.24%	83.13%	83.13%
	PE0							logPE0						
Number of signals	5	5	4	4	4	3	3	5	5	4	4	4	3	2
Number of correct signals	3	4	3	3	3	3	3	3	4	3	3	3	3	2
Proportion of correct signals	60%	80%	75%	75%	75%	100%	100%	60%	80%	75%	75%	75%	100%	100%
Test statistics	0.0449	1.3445	0.6654	0.6654	0.6654	-	-	0.0449	1.3445	0.6654	0.6654	0.6654	-	-
p-value	83.23%	24.62%	41.47%	41.47%	41.47%	-	-	83.23%	24.62%	41.47%	41.47%	41.47%	-	-
	BSEYD10							logBSEYD10						
Number of signals	7	7	6	4	3	4	3	7	4	5	6	5	4	4
Number of correct signals	3	3	2	3	2	3	3	3	3	3	4	3	3	3
Proportion of correct signals	42.86%	42.86%	33.33%	75%	66.67%	75%	100%	42.86%	75%	60%	66.67%	60%	75%	75%
Test statistics	0.436	0.436	1.1741	0.6654	0.1606	0.6654	-	0.436	0.6654	0.0449	0.3212	0.0449	0.6654	0.6654
p-value	50.91%	50.91%	27.86%	41.47%	68.86%	41.47%	-	50.91%	41.47%	83.23%	57.09%	83.23%	41.47%	41.47%
	CAPE10							logCAPE10						
Number of signals	4	3	4	4	3	3	3	4	3	4	3	3	3	2
Number of correct signals	3	3	4	4	3	3	3	3	3	4	3	3	3	2
Proportion of correct signals	75%	100%	100%	100%	100%	100%	100%	75%	100%	100%	100%	100%	100%	100%
Test statistics	0.6654	-	-	-	-	-	-	0.6654	-	-	-	-	-	-
p-value	41.47%	-	-	-	-	-	-	41.47%	-	-	-	-	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 16: **SHCOMP: Accuracy and statistical significance of the prediction models as a function of the confidence level** α. The numbers presented in this table are based on a forecasting horizon $H = 252$ days. With this choice, the uninformed probability that a random guess would correctly identify an equity market downturn is $p_0 = 55.31\%$ Row 1,2 and 3 respectively report the total number of signals generated by the P/E ratio, the number of correct signals, and the proportion of correct signals computed as the ratio of the number of correct signals to the total number of signals. Rows 4 and 5 respectively report the test statistics and p-value for the P/E ratio. The subsequent rows present the same information for the log P/E ratio.

worst performing measures. On aggregate the measures behave as expected: their accuracy increases as α increases, but they are not particularly sensitive to our initial choice $\alpha = 0.95$.

The results in Table 17 indicate that BSEYD0, PE0, logPE0, BSEYD10 and logBSEYD10 perform best at $H = 126$, while CAPE10 and logCAPE10 reach 100% accuracy at $H = 126$. In fact, all the measures except log-BSEYD0 and logBSEYD10 are significant at the 90% confidence with the choice $H = 126$. We conclude that the measures are sensitive to the forecasting horizon, and that the standard choice $H = 252$ is suboptimal on this dataset. This conclusion comes in support of the second hypothesis we suggested to explain the relatively poor performance of the BSEYD models. It does not, however, fully explain this relative underperformance.

On the Shenzhen market, the measures are resilient to a change in the accuracy parameter α, as shown in Table 18. The logBSEYD0, PE0, CAPE10

	63	126	189	252	63	126	189	252
Uninformed probability p_0	22.21%	33.1%	44.2%	55.31%	22.21%	33.1%	44.2%	55.31%
	BSEYD0				logBSEYD0			
Number of signals	4	4	4	4	5	5	5	5
Number of correct signals	2	3	3	3	1	2	2	2
Proportion of correct signals	50%	75%	75%	75%	20%	40%	40%	40%
Test statistics	1.4776	2.9393	1.5663	0.6654	0.0145	0.1043	0.0361	0.4713
p-value	22.41%	8.64%†	21.07%	41.47%	90.41%	74.67%	84.93%	49.24%
	PE0				logPE0			
Number of signals	4	4	4	4	4	4	4	4
Number of correct signals	1	3	3	3	2	3	3	3
Proportion of correct signals	25%	75%	75%	75%	50%	75%	75%	75%
Test statistics	0.0175	2.9393	1.5663	0.6654	1.4776	2.9393	1.5663	0.6654
p-value	89.48%	8.64%†	21.07%	41.47%	22.41%	8.64%†	21.07%	41.47%
	BSEYD10				logBSEYD10			
Number of signals	3	3	3	3	5	5	5	5
Number of correct signals	1	2	2	2	2	3	3	3
Proportion of correct signals	33.33%	66.67%	66.67%	66.67%	40%	60%	60%	60%
Test statistics	0.1947	1.4076	0.6132	0.1606	0.7951	1.5118	0.5019	0.0449
p-value	65.90%	23.55%	43.36%	68.86%	37.26%	21.89%	47.87%	83.23%
	CAPE10				logCAPE10			
Number of signals	3	3	3	3	3	3	3	3
Number of correct signals	2	3	3	3	2	3	3	3
Proportion of correct signals	66.67%	100%	100%	100%	66.67%	100%	100%	100%
Test statistics	2.7014	-	-	-	2.7014	-	-	-
p-value	10.03%	-	-	-	10.03%	-	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 17: **SHCOMP: Accuracy and statistical significance of the BSEYD and log BSEYD as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, the BSEYD ratio generated 4 signals, and the log BSEYD ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.The subsequent rows present the same information for the log P/E ratio.

	0.8	0.85	0.9	0.925	0.95	0.975	0.99	0.8	0.85	0.9	0.925	0.95	0.975	0.99
	BSEYD0							logBSEYD0						
Number of signals	8	6	7	6	6	5	4	5	6	7	8	7	4	4
Number of correct signals	7	6	7	6	6	5	4	5	6	7	8	7	4	4
Proportion of correct signals	87.5%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Test statistics	1.7971	-	-	-	-	-	-	-	-	-	-	-	-	-
p-value	18.01%	-	-	-	-	-	-	-	-	-	-	-	-	-
	PE0							logPE0						
Number of signals	8	9	6	7	6	6	6	8	8	6	6	6	7	4
Number of correct signals	8	9	6	7	6	6	6	7	8	6	6	6	7	4
Proportion of correct signals	100%	100%	100%	100%	100%	100%	100%	87.5%	100%	100%	100%	100%	100%	100%
Test statistics	-	-	-	-	-	-	-	1.7971	-	-	-	-	-	-
p-value	-	-	-	-	-	-	-	18.01%	-	-	-	-	-	-
	BSEYD10							logBSEYD10						
Number of signals	7	6	5	6	7	4	3	10	9	7	8	7	3	3
Number of correct signals	7	6	5	5	6	4	3	10	8	7	8	7	3	3
Proportion of correct signals	100%	100%	100%	83.33%	85.71%	100%	100%	100%	88.89%	100%	100%	100%	100%	100%
Test statistics	-	-	-	0.8162	1.2826	-	-	-	2.3477	-	-	-	-	-
p-value	-	-	-	36.63%	25.74%	-	-	-	12.55%	-	-	-	-	-
	CAPE10							logCAPE10						
Number of signals	8	9	8	9	6	5	4	8	9	6	6	5	4	3
Number of correct signals	8	9	8	9	6	5	4	8	9	6	6	5	4	3
Proportion of correct signals	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%	100%
Test statistics	-	-	-	-	-	-	-	-	-	-	-	-	-	-
p-value	-	-	-	-	-	-	-	-	-	-	-	-	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 18: **SZECOMP: Accuracy and statistical significance of the prediction models as a function of the confidence level** α. The numbers presented in this table are based on a forecasting horizon $H = 252$ days. With this choice, the uninformed probability that a random guess would correctly identify an equity market downturn is $p_0 = 66.99\%$ Row 1,2 and 3 respectively report the total number of signals generated by the P/E ratio, the number of correct signals, and the proportion of correct signals computed as the ratio of the number of correct signals to the total number of signals. Rows 4 and 5 respectively report the test statistics and p-value for the P/E ratio. The subsequent rows present the same information for the log P/E ratio.

and logCAPE10 maintain a 100% accuracy over the entire range of accuracy parameters. BSEYD0 and logPE0 have a 100% accuracy on the range $[0.85, 0.99]$, while BSEYD10 and logBSEYD10 are 100% accurate over most of the range. None of the measures is less than 83.33% accurate.

Seven measures have a 100% accuracy at $H = 189$ and $H = 252$. At a horizon $H = 126$ days, the accuracy of five of the measures is statistically significant at the 95% confidence level. At this horizon, the accuracy of the worst performing measures is 71.43%, far above the prior probability $p_0 = 44\%$. Further reducing the time horizon to $H = 63$ days, reduces the accuracy of the measures. Still, three of the eight measure are statistically significant at the 90% confidence level.

	63	126	189	252	63	126	189	252
Uninformed probability p_0	24.93%	44.22%	58.47%	66.99%	24.93%	44.22%	58.47%	66.99%
	BSEYD0				logBSEYD0			
Number of signals	6	6	6	6	7	7	7	7
Number of correct signals	3	5	6	6	4	6	7	7
Proportion of correct signals	50%	83.33%	100%	100%	57.14%	85.71%	100%	100%
Test statistics	1.7367	3.9209	-	-	3.2717	5.2181	-	-
p-value	18.76%	4.77%*	-	-	7.05%†	2.24%*	-	-
	PE0				logPE0			
Number of signals	6	6	6	6	6	6	6	6
Number of correct signals	2	5	6	6	2	5	6	6
Proportion of correct signals	33.33%	83.33%	100%	100%	33.33%	83.33%	100%	100%
Test statistics	0.212	3.9209	-	-	0.212	3.9209	-	-
p-value	64.52%	4.77%*	-	-	64.52%	4.77%*	-	-
	BSEYD10				logBSEYD10			
Number of signals	7	7	7	7	7	7	7	7
Number of correct signals	4	5	6	6	4	5	7	7
Proportion of correct signals	57.14%	71.43%	85.71%	85.71%	57.14%	71.43%	100%	100%
Test statistics	3.2717	2.1193	2.4553	1.2826	3.2717	2.1193	-	-
p-value	7.05%†	14.54%	11.71%	25.74%	7.05%†	14.54%	-	-
	CAPE10				logCAPE10			
Number of signals	6	6	6	6	5	5	5	5
Number of correct signals	3	5	6	6	2	4	5	5
Proportion of correct signals	50%	83.33%	100%	100%	40%	80%	100%	100%
Test statistics	1.7367	3.9209	-	-	0.5465	2.6916	-	-
p-value	18.76%	4.77%*	-	-	45.98%	10.09%	-	-

† significant at the 10% level;
* significant at the 5% level;
** significant at the 1% level;
*** significant at the 0.5% level.

Table 19: **SZECOMP: Accuracy and statistical significance of the BSEYD and log BSEYD as a function of the forecasting horizon H.** The numbers presented in this table are based on a confidence parameter $\alpha = 0.95$. With this choice, the BSEYD ratio generated 4 signals, and the log BSEYD ratio produced 18 signals. Row 1 presents the uninformed probability p_0 that a random guess would correctly identify an equity market downturn. Row 3 reports the number of correct signals, row 4, the proportion of correct signals as the ratio of the number of correct signals to the total number of signals for the P/E ratio. Rows 5 and 6 respectively report the test statistics and p-value.The subsequent rows present the same information for the log P/E ratio.

7 Can We Use the Predictors to Construct Active Investment Strategies?

We tested the statistical significance of the P/E ratio, CAPE, and BSEYD as predictors of equity market downturns in China. The next question is whether these crash predictors are relevant to asset managers. Can we use the P/E ratio, the CAPE, and the BSEYD to construct active investment strategies capable of outperforming a simple buy and hold investment?

To construct active strategies, we need to generate "buy" and "sell" signals based on the predictors and then trade in and out of the index based on these signals. Then, the performance of the active strategies is assessed against a simple buy and hold investment.

7.1 Buy and Sell Signals

We use the P/E ratio, CAPE, and BSEYD and their logarithms to generate sell and buy signals: a sell signal occurs whenever a measure (PE0, BSEYD0, CAPE10, BSEYD10) crosses above a threshold, and a buy signal occurs whenever the same measure crosses below a threshold:

$$\text{Signal}_{\text{Sell}}(t) = M(t) - K_{\text{Sell}}(t) > 0, \qquad \text{Signal}_{\text{Buy}}(t) = M(t) - K_{\text{Buy}}(t) < 0$$

All active strategies start with a 100% investment in the SHCOMP index or SZECOMP index. When a sell signal occurs, the portfolio manager gradually sells the position in the index and invests the proceeds overnight at the call money/interbank rate for China (source: Federal Reserve Bank of St. Louis, `https://fred.stlouisfed.org`), which we use as a proxy for the short-term rate. When a buy signal occurs, the portfolio manager gradually invests in the stock index until the portfolio is fully invested. The active strategies can neither be short on the index, nor leveraged on the index, so at all time the portfolio is invested between 0% and 100% in the index.

We define the time-varying thresholds as a dynamic confidence interval based on a Normal distribution:

$$K_{\text{Sell}} = \mu_t^h + \alpha_{\text{Sell}} \times \sigma_t^h, \qquad K_{\text{Buy}} = \mu_t^h - \alpha_{\text{Buy}} \times \sigma_t^h$$

The primary determinants of the thresholds are the reliability factors α_{Sell} and α_{Buy}, or equivalently, the confidence levels c_{Sell} and c_{Buy}. In practice, the confidence level for crash prediction is usually 95%, but Berge et al. (2008)

suggest using a 90%-95% confidence for sell thresholds and an 80%-85% confidence for buy thresholds.

In this paper, we determine the confidence levels $(c_{\text{Sell}}, c_{\text{Buy}}) \in (0,1)^2$ by optimization to maximize the Sortino ratio of the active strategies. Because the optimization problem is nonlinear and nonsmooth, we use the evolutionary method in Frontline System's Solver for Microsoft Excel, with a population of 200. Since nonlinear optimization methods are at risk of converging to a local optimum, we performe the optimization starting from different initial values and check the stability of the solutions obtained.

7.2 Trading on the Signals

Buy and sell transactions consider two sources of market liquidity constraints: cost, and trading window. The transaction cost is the amount paid as commissions or bid-ask spread to enter or exit a position. The trading window represents the number of days to execute the trade fully. Typically, the bigger the position, the higher the transaction cost and the longer the trading window.

For a given position size, transaction costs and trading window are inversely related. A longer trading window typically provides the asset manager with more time to trade, helping to reduce the effective transaction costs and price impact of the trade. However, a longer trading window increases the opportunity cost, in our case the cost associated with exiting the market too late to protect the portfolio from a downturn, and the cost of entering the market too later to take advantage of a recovery.

While the effect of transaction costs on active strategies is straightforward as higher transaction costs reduce returns, the impact of the trading window is less obvious: will a change in the trading window increase the opportunity cost materially? To answer this question, we test the investment strategies on the base case of a trading window of 90 days, and on two further cases: 30 days and 120 days.

Regarding execution, we assume for simplicity that the portfolio manager sells a constant fraction. In the base case, the manager would sell (buy) 1/90th of their holdings each day of the 90-day trading window, starting on the first day following a sell (buy) signal, and up until either:

1. The manager has shifted the entire portfolio into the short-term rate

(index) at the end of the 90-day trading window; or

2. A buy (sell) signal occurs, reversing the trading flow. Hence, we do not assume any particular trading behavior during the window, although in practice portfolio managers will try to optimize their execution to add value to the strategy.

Note that the buy and hold investment is not subject to liquidity considerations: it remains fully invested in the index.

7.3 Performance Evaluation

In total, we construct 20 active strategies for Shanghai and Shenzhen:

- Full period (1990-2016 for SHCOMP, 1991-2016 for SZECOMP): PE0 and logPE0;
- Period from October 31, 2006 to June 30, 2016: PE0 and logPE0, BSEYD0 and logBSEYD0, CAPE10 and logCAPE10, BSEYD10 and logBSEYD10.

The reference point for the evaluation is a simple buy and hold investment, corresponding to a constant 100% position in the SHCOMP or SZECOMP index.

For each strategy, we compute the daily excess return over the short-term rate, net of transaction costs. Excess returns absorb the stochasticity of the short-term rate and provide a more accurate measurement of the mean-risk ratios and risk-adjusted measures than nominal returns.

We assess the performance of active strategies against a simple buy and hold strategy using four categories of metrics: descriptive statistics, risk measures, mean-to-risk ratios, and active risk and returns. The descriptive statistics include mean, standard deviation, skewness, and kurtosis. The risk measures include semi-deviation, maximum daily drawdown, 95% daily Value-at-Risk (VaR) and Conditional Value-at-Risk (CVaR), and 99% VaR and CVaR.

We then use these risk measures to define several mean-to-risk ratios: Sharpe, Sortino, mean-to-95%VaR, mean-to-95%CVaR, mean-to-99%VaR, and mean-to-99%CVaR. The mean-to-VaR and mean-to-CVaR are expressed in percentage excess returns and are computed about the mean.

Finally, we produced active risk and return statistics by regressing the excess returns of the active strategies against the excess return of the buy and hold investment. The slope of the regression gives the beta of the active strategy, the intercept corresponds to the risk-adjusted return or Jensen's alpha, the standard error of the regression is the active risk or omega. The ratio of Jensen's alpha to omega is the appraisal ratio.

7.4 Results: Full Period

Table 20 presents the optimal confidence levels and the four categories of metrics (descriptive statistics, risk measures, mean-to-risk ratios and active risk and returns) for the buy and hold, PE0 and logPE0 strategies on the Shanghai and Shenzhen stock markets. For the active strategies, the percentage change in the statistics over the buy and hold investment are computed. Overall, we find that it is possible to construct active PE0 and logPE0-based strategies outperforming a simple buy and hold investment.

7.4.1 Results on the Shanghai Market

On the Shanghai market, active strategies outperform a buy and hold investment on all metrics. The confidence levels are optimized at 89%-90% for sell signals and 83%-85% for buy signals.

For the descriptive statistics, active strategies generate a higher return than the index (3 to 5 basis points a day, about 7% to 13% returns per year), while reducing risk by 20% to 25%. Active strategies have a more pronounced right skew than the index, indicating that active strategies have much more instances of returns above than below the mean. The kurtosis also increases noticeably, implying a hollowing out of the bulk of the distribution, and fatter tails.

The risk measures indicate that the risk of the active strategies is lower than that of the index. For example, the 99% CVaR is around 2% lower than that of the index, meaning that for the 1% worst daily loss incurred over the period, the active strategies would have lost 2% fewer assets on average per day than the index. However, the maximum drawdown is not much lower than that of the index, implying that the active strategies were unable to avoid all of the worst one-day losses.

All of the mean-to-risk ratios show a dramatic improvement over the index (56% to 226%). The Sortino ratio improved the most because it is the objective function in our optimization. All the other ratios improved over the index by 56% to 190%, meaning that the active strategies deliver more excess return per unit of risk (as measured by the standard deviation, semi-deviation, VaRs, and CVaRs) than the index.

Considering active risk and return, the beta of the active strategies is only 0.5 to 0.6. Hence the active strategies are defensive. Both active strategies generate positive alpha (between 0.6 basis points and 2.75 basis points a day, roughly 4% to 7% per year), but with a substantial active risk as indicated by omega, resulting in a moderately positive appraisal ratio.

7.4.2 Results on the Shenzhen Market

The results on the Shenzhen market are less clear-cut overall, although both active strategies outperformed the index. The optimal confidence for the PE0 and logPE0 strategy are 23% and 43% respectively for sell decisions, and 99% and 93% respectively for buy decision.

	SHCOMP			SZECOMP		
	Buy and Hold	PE0	logPE0	Buy and Hold	PE0	logPE0
Confidence Parameter						
Sell		89.21%	90.24%		22.96%	42.95%
Buy		82.98%	85.06%		99.27%	93.04%
Descriptive Statistics						
Mean Return	0.12%	0.17% (47%)	0.15% (28%)	0.04%	0.06% (63%)	0.04% (-2%)
Standard deviation	3.29%	2.43% (-26%)	2.68% (-19%)	1.9%	0.97% (-49%)	1.02% (-47%)
Skewness	5.45	13.31 (144%)	10.19 (87%)	-0.38	-0.85 (123%)	0.19 (-149%)
Kurtosis	111.29	360.33 (224%)	248.29 (123%)	5.99	13.49 (125%)	19.54 (226%)
Risk Measures (From Mean)						
Semi-deviation	2.16%	0.97% (-55%)	1.29% (-40%)	2.06%	0.85% (-59%)	0.82% (-60%)
Maximum drawdown from mean	17.91%	13.55% (-24%)	16.73% (-7%)	8.93%	8.64% (-3%)	8.93% (0%)
VaR 95	3.17%	1.62% (-49%)	2.15% (-32%)	3.13%	1.98% (-37%)	1.12% (-64%)
CVaR 95	5.37%	3.17% (-41%)	3.8% (-29%)	4.81%	3.17% (-34%)	2.14% (-56%)
VaR 99	6.73%	4.12% (-38%)	4.76% (-29%)	6.16%	4.53% (-27%)	2.81% (-54%)
CVaR 99	8.85%	6.33% (-28%)	6.97% (-21%)	7.22%	6.07% (-16%)	4.12% (-43%)
Mean-Risk Ratios						
Sharpe ratio	0.036	0.0718 (99%)	0.0564 (57%)	0.0191	0.0611 (220%)	0.035 (83%)
Sortino ratio	0.055	0.1797 (227%)	0.1176 (114%)	0.0177	0.0701 (296%)	0.0434 (146%)
Mean-to-VaR95	0.0375	0.1078 (188%)	0.0705 (88%)	0.0116	0.0299 (158%)	0.0317 (174%)
Mean-to-CVaR95	0.0221	0.055 (149%)	0.0398 (80%)	0.0075	0.0187 (148%)	0.0167 (121%)
Mean-to-VaR99	0.0176	0.0423 (140%)	0.0318 (80%)	0.0059	0.0131 (122%)	0.0127 (115%)
Mean-to-CVaR99	0.0134	0.0275 (105%)	0.0217 (62%)	0.005	0.0098 (94%)	0.0086 (72%)
Active Risk and Return						
Beta	1 (0)	0.49 (0.0062)	0.64 (0.0058)	1 (0)	0.48 (0.0075)	0.21 (0.0064)
Alpha	0% (0)	0.0275% (0.0001)	0.0062% (0.0001)	0% (0)	0.0279% (0.0001)	0.0156% (0.0001)
Omega	0%	1.1629%	1.0998%	0	0.8525%	0.7223%
Appraisal ratio	0	0.0237	0.0057	0	0.0327	0.0215

Table 20: **Optimal Active Strategies (Full Period)**. For each investment strategy considered in column, the table presents the buy and sell confidence parameters (for active strategies), descriptive statistics, risk measures, mean-risk ratios and active risk and return statistics. Percentages between brackets next to the descriptive statistics, risk measures and mean-risk ratios indicate the percentage change compared to the buy and hold investment. Numbers between brackets next to the active risk and return statistics indicate the standard error of estimates.

For the descriptive statistics, the mean excess return of the PE0-based strategy is markedly higher than that of the index, while the excess return of the logPE0-based strategy is virtually identical to that of the index. Both strategies achieve a similar reduction in the standard deviation, but they diverge again slightly on the skewness: the PE0-based strategy is a bit more skewed to the left than the index, while the logPE0 strategy is somewhat skewed to the right. Both strategies have much fatter tails than the index.

The risk measures indicate that the risk of the active strategies is lower than that of the index, although the maximum drawdown remains close to that of the index. As a result, the mean-to-risk ratios are all markedly higher than that of the index. The improvement ranges from 71% for the mean-to-CVaR99 of the logPE0 strategy to 296% for the Sortino ratio of the PE0 strategy.

The beta of the active strategies is significantly lower than 1. Both strategies generate positive alpha (between 1.56 basis points and 2.70 basis points a day, roughly 1.5% to 7% per year), but with a large active risk as measured by omega, resulting in a moderately positive appraisal ratio.

7.5 Results: Period 2006-2016

Similar to Table 20 , Tables 21 and 22 present, for the Shanghai and Shenzhen stock market respectively, the optimal confidence levels and four categories of metrics for the buy and hold, PE0, logPE0, BSEYD0, logBSEYD0, CAPE10, logCAPE10, BSEYD10 and logBSEYD10 strategies over the October 31, 2006 to June 30, 2016 period. For the active strategies, the percentage change in the statistics over the buy and hold investment are computed. Overall, we find that it is possible to construct active strategies outperforming a simple buy and hold investment.

7.5.1 Results on the Shanghai Market

The confidence parameters displayed in Tables 21 appear relatively uniform, although the optimal confidence levels for the sell and buy decisions are asymmetric. The optimal confidence levels are between 40% and 76% for sell decisions and between 0.5% and 5% for buy decisions. These results suggest that a portfolio manager would only need to be 40% to 76% confident that a correction will occur to sell, and only 0.5% to 5% confident that an increase will happen to buy.

For the descriptive statistics, the active strategies do not outperform the index regarding mean excess return, but they generate between 60% and 90% less risk than the index as measured by the standard deviation of excess returns. The return distribution of the active strategies is less skewed to the right than that of the index and has fatter tails.

All the risk measures related to the active strategies are between 51% and 99% lower than for the index, confirming our earlier observation on the standard deviation. Even the maximum drawdown dropped by 51% to 75%, from 9.26% to a range of 2.15% to 4.44%, which was not the case over the full period.

As a result, we observe a dramatic increase in all the mean-to-risk ratios, and in particular the mean-to-VaR and mean-to-CVaR ratios. This observation is consistent with the interpretation that the optimal strategies on the SHCOMP reduce risk more than they increase returns.

The analysis of active risk and return provides further evidence in this direction: the beta of all the strategies is not significantly different from 0, while all the strategies generate moderate amount of alpha (between 1.55 basis points and 2.12 basis points a day, roughly 1.5% to 5.3% per year), with a moderate omega, resulting in a relatively high appraisal ratio.

7.5.2 Results on the Shenzhen Market

The results on the Shenzhen market are more heterogeneous than on the Shanghai market, as was the case in the full period. For six out of eight strategies, the optimal confidence level for sell decisions is between 43% and 54%, and between 26% and 30% for the remaining two strategies. However, the confidence levels for buy decisions show more variability, from lows of 5% to highs of 59%.

For the descriptive statistics, we observe that the mean excess return of the active strategies is lower than the index. The standard deviation for the active strategies is between 46% and 78% lower than for the index, while the skewness is consistent with that of the index, and the kurtosis is much higher than that of the index.

All the active strategies outperform the index in terms of risk measures. Active strategies are between 25% and 88% less risky than the index, except for the maximum drawdown where two of the strategies do not improve markedly on the index.

The mean-to-risk ratios of the active strategies outperform the index markedly, especially the Sharpe ratio, Sortino ratio and mean-to-VaR95 ratio.

Regarding active risk and return, the active strategies have low betas, ranging from 0.08 to 0.34. They generate a reasonably large alpha (between 2.2 basis points and 5 basis points a day, about 5.5% to 12.6% per year), with a sizable omega, resulting in a relatively high appraisal ratio.

	Buy and Hold	PE0	logPE0	BSEYD0	logBSEYD0	CAPE10	logCAPE10	BSEYD10	logBSEYD10
Confidence parameter									
Sell		40.18%	53.92%	60.77%	41.72%	46.73%	49.79%	75.84%	75.84%
Buy		0.50%	0.74%	5.34%	4.78%	2.88%	2.49%	0.50%	0.50%
Descriptive statistics									
Mean Return	0.02%	0.02% (8%)	0.02% (7%)	0.02% (-13%)	0.02% (-6%)	0.02% (-15%)	0.02% (-13%)	0.02% (3%)	0.04% (104%)
Standard deviation	1.87%	0.24% (-87%)	0.24% (-87%)	0.42% (-77%)	0.36% (-81%)	0.38% (-80%)	0.37% (-80%)	0.22% (-88%)	0.72% (-62%)
Skewness	-0.47	5.56 (-1294%)	5.57 (-1294%)	-0.35 (-25%)	0.68 (-246%)	-0.04 (-92%)	0.06 (-112%)	5.93 (-1371%)	-1.09 (133%)
Kurt	6.38	119.14 (1768%)	119.13 (1768%)	44.97 (605%)	49.77 (680%)	48.88 (667%)	50.29 (689%)	121.06 (1798%)	31.69 (397%)
Risk Measures (From Mean)									
Semi-deviation	2.03%	0.13% (-94%)	0.13% (-94%)	0.31% (-85%)	0.25% (-88%)	0.27% (-87%)	0.26% (-87%)	0.12% (-94%)	0.56% (-72%)
Maximum drawdown from mean	9.26%	2.28% (-75%)	2.28% (-75%)	4.58% (-51%)	4.04% (-56%)	4.44% (-52%)	4.44% (-52%)	2.15% (-77%)	7.31% (-21%)
VaR 95	3.07%	0.01% (-100%)	0.01% (-100%)	0.01% (-100%)	0.01% (-100%)	0.01% (-100%)	0.01% (-100%)	0.01% (-100%)	0.64% (-79%)
CVaR 95	4.84%	0.04% (-99%)	0.04% (-99%)	0.11% (-98%)	0.08% (-98%)	0.1% (-98%)	0.09% (-98%)	0.04% (-99%)	1.82% (-62%)
VaR 99	5.97%	0.18% (-97%)	0.18% (-97%)	1.54% (-74%)	0.01% (-100%)	1.39% (-77%)	1.35% (-77%)	0.17% (-97%)	2.38% (-60%)
CVaR 99	7.29%	1.05% (-86%)	1.05% (-86%)	2.51% (-66%)	0.08% (-99%)	2.19% (-70%)	2.14% (-71%)	0.95% (-87%)	3.93% (-46%)
Mean-Risk Measures									
Sharpe ratio	0.0108	0.0906 (742%)	0.0898 (735%)	0.0412 (283%)	0.0521 (385%)	0.0455 (323%)	0.0475 (341%)	0.093 (765%)	0.0573 (433%)
Sortino ratio	0.0099	0.1696 (1,604%)	0.1682 (1591%)	0.0562 (465%)	0.0757 (661%)	0.064 (543%)	0.0672 (575%)	0.1781 (1,690%)	0.0735 (639%)
Mean-to-VaR95	0.0066	1.5865 (24,048%)	1.5943 (24,166%)	1.8549 (28,133%)	1.7417 (26,410%)	1.8922 (28,700%)	1.8524 (28,094%)	1.6388 (24,844%)	0.0644 (881%)
Mean-to-CVaR95	0.0042	0.5274 (12,560%)	0.5473 (13,037%)	0.1526 (3,563%)	0.2357 (5,557%)	0.1784 (4,183%)	0.1881 (4,416%)	0.5593 (13,327%)	0.0227 (444%)
Mean-to-VaR99	0.0034	0.1241 (3,573%)	0.1232 (3,547%)	0.0114 (236%)	1.7417 (51,465%)	0.0123 (265%)	0.0129 (283%)	0.1242 (3,576%)	0.0173 (413%)
Mean-to-CVaR99	0.0028	0.0207 (650%)	0.0206 (644%)	0.007 (152%)	0.2357 (8,420%)	0.0078 (183%)	0.0082 (196%)	0.0218 (687%)	0.0105 (279%)
Active Risk and Return									
Beta	1 (0)	0.03 (0.0026)	0.03 (0.0026)	0.09 (0.0043)	0.07 (0.0038)	0.08 (0.0039)	0.08 (0.0038)	0.03 (0.0024)	0.19 (0.007)
Alpha	0% (0)	0.0212% (0)	0.021% (0)	0.0156% (0.0001)	0.0174% (0.0001)	0.0155% (0.0001)	0.0159% (0.0001)	0.0201% (0)	0.0374% (0.0001)
Omega	0	0.2345%	0.2345%	0.3869%	0.3351%	0.3444%	0.3392%	0.2167%	0.6245%
Appraisal ratio	0	0.0905	0.0897	0.0403	0.052	0.045	0.047	0.093	0.0598

Table 21: **Optimal Active Strategies on the SHCOMP (2006-2106)**. For each investment strategy considered in column, the table presents the buy and sell confidence parameters (for active strategies), descriptive statistics, risk measures, mean-risk ratios and active risk and return statistics. Percentages between brackets next to the descriptive statistics, risk measures and mean-risk ratios indicate the percentage change compared to the buy and hold investment. Numbers between brackets next to the active risk and return statistics indicate the standard error of estimates.

7.6 Discussion

These results suggest that it is possible to use PE and BSEYD-based measures to construct active strategies that outperform a simple buy and hold investment on a risk-adjusted basis on the Shanghai and Shenzhen equity markets. The conclusions are stronger on the Shanghai equity market than on the Shenzhen equity market.

The confidence levels provide the most striking observation of this study. The confidence levels for buy and sell do not match, and they are not close to each other. For sell decisions, the confidence level is typically much lower for asset management applications than for statistical investigation. This means that asset managers may make wrong decisions occasionally but still outperform a strategy that only trades when the confidence level is high.

So far, these active strategies are all based on a 90-day trading window, meaning that managers have 90 days to trade following a signal. We also performed the same analysis on trading windows of 30 days and 120 days. Overall, the strategies performed better with longer trading windows.

At first, this result appears counterintuitive: experience suggests that longer trading windows should result in higher opportunity costs and therefore reduce the effectiveness of active strategies. However, this statement makes an implicit assumption: that the trading signal received should prompt immediate action. This is not the case with the PE, CAPE, and BSEYD. These measures tend to peak and generate a signal six to nine months before a downturn occurs (as evidenced by the robustness tests).

This early warning feature gives the PE, CAPE, and BSEYD an advantage from an asset management perspective because it leaves portfolio managers ample time to shift their positions in an orderly manner, and at minimum liquidity cost to their funds. Furthermore, this early warning feature explains why acting immediately on the signal may be a bad idea: by selling too soon (or buying too early), a portfolio manager may miss out a part of the rally (or invest while the market is still correcting).

	Buy and Hold	PE0	logPE0	BSEYD0	logBSEYD0	CAPE10	logCAPE10	BSEYD10	logBSEYD10
Confidence parameter									
Sell		50.17%	54.02%	42.85%	30.31%	47.71%	53.01%	25.83%	43.00%
Buy		54.68%	59.2%	4.7%	10.99%	26.46%	29.99%	6.99%	7.93%
Descriptive statistics									
Mean Return	0.07%	0.04% (-36%)	0.07% (10%)	0.03% (-52%)	0.05% (-16%)	0.04% (-46%)	0.04% (-42%)	0.04% (-32%)	0.03% (-57%)
Standard deviation	2.1%	0.78% (-63%)	1.14% (-46%)	0.66% (-69%)	0.82% (-61%)	0.6% (-71%)	0.46% (-78%)	0.6% (-72%)	0.49% (-76%)
Skewness	-0.67	-0.83 (23%)	-1.06 (58%)	-0.98 (45%)	-1.72 (156%)	-0.81 (20%)	-0.35 (-48%)	-1.21 (79%)	-0.9 (34%)
Kurt	5.20	14.43 (177%)	12.65 (143%)	18.82 (262%)	31.7 (509%)	18.79 (261%)	20.9 (302%)	31.34 (502%)	32.64 (527%)
Risk Measures (From Mean)									
Semi-deviation	2.39%	0.67% (-72%)	1.03% (-57%)	0.53% (-78%)	0.64% (-73%)	0.48% (-80%)	0.34% (-86%)	0.46% (-81%)	0.38% (-84%)
Maximum drawdown from mean	8.94%	6.31% (-29%)	8.94% (0%)	5.11% (-43%)	8.93% (0%)	5.6% (-37%)	3.56% (-60%)	7.16% (-20%)	4.9% (-45%)
VaR 95	3.75%	1.19% (-68%)	1.73% (-54%)	0.78% (-79%)	0.45% (-88%)	0.65% (-83%)	0.41% (-89%)	0.44% (-88%)	0.28% (-93%)
CVaR 95	5.52%	2.17% (-61%)	3.15% (-43%)	1.86% (-66%)	1.98% (-64%)	1.66% (-70%)	1.18% (-79%)	1.51% (-73%)	1.24% (-78%)
VaR 99	6.85%	2.64% (-61%)	3.97% (-42%)	2.45% (-64%)	2.89% (-58%)	2.18% (-68%)	1.82% (-73%)	2.18% (-68%)	1.9% (-72%)
CVaR 99	7.71%	3.78% (-51%)	5.74% (-26%)	3.5% (-55%)	4.97% (-36%)	3.11% (-60%)	2.43% (-68%)	3.43% (-56%)	2.88% (-63%)
Mean-Risk Measures									
Sharpe ratio	0.0311	0.0532 (71%)	0.0634 (104%)	0.0479 (54%)	0.0669 (115%)	0.0588 (89%)	0.083 (167%)	0.0746 (140%)	0.0566 (82%)
Sortino ratio	0.0273	0.0624 (128%)	0.0701 (156%)	0.0594 (117%)	0.0859 (214%)	0.0738 (170%)	0.1104 (304%)	0.0964 (252%)	0.0738 (170%)
Mean-to-VaR95	0.0174	0.0349 (100%)	0.0416 (139%)	0.0404 (132%)	0.1211 (596%)	0.0541 (211%)	0.0923 (431%)	0.1004 (477%)	0.1004 (477%)
Mean-to-CVaR95	0.0118	0.0192 (62%)	0.0229 (94%)	0.017 (43%)	0.0278 (135%)	0.0212 (79%)	0.0321 (172%)	0.0295 (150%)	0.0226 (91%)
Mean-to-VaR99	0.0095	0.0157 (65%)	0.0182 (91%)	0.0129 (35%)	0.019 (99%)	0.0162 (70%)	0.0209 (119%)	0.0205 (115%)	0.0148 (55%)
Mean-to-CVaR99	0.0085	0.011 (30%)	0.0126 (49%)	0.009 (6%)	0.011 (30%)	0.0113 (34%)	0.0156 (84%)	0.013 (54%)	0.0097 (15%)
Active Risk and Return									
Beta	1 (0)	0.18 (0.0069)	0.34 (0.0089)	0.14 (0.0058)	0.17 (0.0074)	0.12 (0.0054)	0.08 (0.0043)	0.13 (0.0053)	0.1 (0.0045)
Alpha	0% (0)	0.03% (0.0001)	0.0501% (0.0002)	0.0221% (0.0001)	0.044% (0.0002)	0.0271% (0.0001)	0.0329% (0.0001)	0.0361% (0.0001)	0.0215% (0.0001)
Omega	0%	0.6869%	0.8903%	0.5835%	0.7419%	0.539%	0.4278%	0.531%	0.4483%
Appraisal ratio	0	0.0437	0.0563	0.0378	0.0594	0.0503	0.077	0.0679	0.048

Table 22: **Optimal Active Strategies on the SZECOMP (2006-2106)**. For each investment strategy considered in column, the table presents the buy and sell confidence parameters (for active strategies), descriptive statistics, risk measures, mean-risk ratios and active risk and return statistics. Percentages between brackets next to the descriptive statistics, risk measures and mean-risk ratios indicate the percentage change compared to the buy and hold investment. Numbers between brackets next to the active risk and return statistics indicate the standard error of estimates.

8 Conclusion And Summary of the Main Results

The Chinese stock market is certainly one of the most interesting and complex equity markets in the world. Its size, scope, structure and the the rapidity of its evolution make it a uniquely challenging, and valuable testing ground for crash prediction models. Overall, the studies in this paper support the application of crash prediction models to the Chinese market, and reveals further research questions both on the behavior of Chinese equity markets, and on crash prediction models.

Over the entire length of the study (1990-2016 for the SHCOMP and 1991-2016 for the SZECOMP), our statistical test rejected the null hypothesis that the logarithm of the P/E do not have predictive power. Moreover, these results are not overly sensitive to changes in the two key parameters of the model: the confidence level α and the forecasting horizon H.

A comparison of the BSEYD, PE and CAPE and their logarithm over a shorter 9-year period, is less conclusive. On the SHCOMP, measures based on the BSEYD do not perform as well as measures based on the P/E and in particular, the CAPE. This is a puzzle because the BSEYD contains more information than the P/E and has been more successful in other markets since 1988. However, all measures perform surprisingly well on the SZECOMP. Two possible explanations for this situation are that (i) the sample is small so any correct or incorrect prediction has a large impact on the accuracy of the measure and its statistical test, and (ii) the market microstructure of the SHCOMP and SZECOMP differ because the Shanghai and Shenzhen stock exchanges were created for two different types of companies: public companies in Shanghai and privately-owned companies in Shenzhen. Both explanations open up avenues for further research.

We also found that it is possible to use the PE, CAPE and BSEYD, and their logarithms, to construct active strategies capable of outperforming a simple buy and hold investment, even when transaction costs and liquidity constraints are factored in. Overall, the active strategies carry a fraction of the risk of the buy and hold investment, and produce higher returns per unit of risk. This observation holds across multiple definitions of risk: standard deviation, semi-deviation, VaR, CVaR, beta, and omega. These findings suggest that PE, CAPE, and BSEYD-based strategies have their place in the asset management world.

9 Acknowledgements

The first author gratefully acknowledges support from Région Champagne Ardennes and the European Union through the RiskPerform Grant. We thank the participants to the 34th International Conference of the French Finance Association and the 10th International Risk Management Conference for their comments and suggestions.

A Appendix: An Overview of Hidden Markov Models

A.1 Basic Structure

We start from the assumption that at any point in time t, the financial index, whether the SHCOMP or the SZECOMP, can be in any of N distinct states $S_1, S_2, \ldots, S_n$. Denote by q_t the actual state of the system at time $t = 1, 2, \ldots$, and by $\{q_t = S_i\}$ the event 'being in state i at time t'.

On any day, the index may change state with probability

$$P\left[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \ldots\right]$$

We further assume that the state transitions satisfy the Markov property, which implies that

$$P\left[q_t = S_j | q_{t-1} = S_i, q_{t-2} = S_k, \ldots\right] = P\left[q_t = S_j | q_{t-1} = S_i\right]$$

we model this as a discrete (first order) Markov Chain with a transition probability matrix $A = (a_{ij})$, $i, j = 1, \ldots, N$ of the form

$$a_{ij} = P\left[q_t = S_j | q_{t-1} = S_i\right] \tag{A.1}$$

where the state transition coefficients a_{ij} satisfy

$$\begin{aligned} a_{ij} &\geq 0 \\ \sum_{j=1}^{N} a_{ij} &= 1 \end{aligned}$$

Next, denote the initial state probabilities by

$$\pi_i = P\left[q_1 = S_i\right], \qquad i = 1, \ldots, N$$

The observation sequence $O = \{O_1, O_2 \ldots O_T\}$ records the actual states that have occurred from time 1 to time T. We allow $O_1, O_2, \ldots$ to be the vector of returns on the index.

In simple cases, we could read the state directly (as an example, we could think about weather condition outside our window). If we have observed state S_1 at time 1, S_3 at time 2, S_1 at time 3 and S_2 at time 4, then the observation sequence is $O = \{S_1, S_3, S_1, S_2\}$. Often, the current state of the system is not directly observable. In this sense the actual sate of the Markov chain is 'hidden'. As a result, we need to rely on observations to infer the current state of the market. For example, the current state of the a financial market is not directly observable: we need to infer it from the returns we observe.

The theory of HMM was originally built around the idea of discrete *observation symbols* associated with each states. These observation symbols form an alphabet of size M. We denote by V the set of all observation symbols, i.e.

$$V = \{v_1, v_2, \ldots, v_k, \ldots v_M\}$$

Think of V as a set of letters or sounds (music notes, syllables...)

The probability of the observation symbol given that the system is in state j is

$$b_j(k) = P\left[v_k \text{ at } t | q_t = S_j\right], \qquad 1 \leq j \leq N, 1 \leq k \leq M \tag{A.2}$$

The probability distribution of the observation symbol is the matrix $B = (b_j(k))$, $1 \leq j \leq N, 1 \leq k \leq M$.

The idea of a discrete observation set does work for simple coin toss or ball-and-urn experiments as well as for some data processing applications, but it has severe limitations for financial markets where the observation sequence is represented by asset returns.

As a result, we need to change the standard model to allow continuous observation sets and continuous probability distributions.

To that effect, we model the returns in each state as a M-component Gaussian mixture. The mathematical specification of this model is:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm} \mathcal{N}\left(\mathbf{O}, \mu_{jm}, \mathbf{\Sigma}_{jm}\right), \quad 1 \leq j \leq N \tag{A.3}$$

where

- $\mathbf{O}$ is a d-dimensional observation vector;

- $\mathcal{N}$ is the Gaussian pdf[4].
- c_{jm} is the mixture coefficient for the j-th state and m-th mixture;
- μ_{jm} is the mean vector for the j-th state and m-th mixture;
- $\mathbf{\Sigma}_{jm}$ is the covariance matrix for the j-th state and m-th mixture.

The mixture coefficient c_{jm} satisfies the following constraints:

$$\begin{aligned} c_{jm} &\geq 0 \\ \sum_{m=1}^{M} c_{jm} &= 1 \end{aligned} \quad \text{(A.4)}$$

Moreover, for b to be a properly defined pdf we need to have

$$\int_{-\infty}^{+\infty} b_j(x)dx = 1, \quad 1 \leq j \leq N \quad \text{(A.5)}$$

When $M = 1$, we revert to the case where the returns in each state are conditionally jointly-Normally distributed.

To sum things up, our HMM model is comprised of:

1. N unobervable states $S_1, S_2, \ldots, S_N$;
2. a $N \times N$ transition probability matrix $A = (a_{ij})$ where

$$a_{ij} = P\left[q_t = S_j | q_{t-1} = S_i\right] \quad \text{(A.6)}$$

3. initial state probabilities

$$\pi_i = P\left[q_1 = S_i\right], \qquad i = 1, \ldots, N$$

4. a sequence of d-dimensional vectors $\{\mathbf{O}_t\}_{t=1,\ldots}$ with observation probability given by a M-dimensional Gaussian mixture:

$$b_j(\mathbf{O}) = \sum_{m=1}^{M} c_{jm} \mathcal{N}\left(\mathbf{O}, \mu_{jm}, \mathbf{\Sigma}_{jm}\right), \quad 1 \leq j \leq N \quad \text{(A.7)}$$

We denote by $B(\mathbf{O})$ the N-dimensional pdf vector.

[4]Any log-concave or elliptically-symmetric probability would work, although in reality most people will use Gaussian distributions

A.2 The Three (Basic) Problems for HMMs And How to Solve Them

We express the set of model parameters λ as $\lambda = (A, B, \pi)$. Rabiner (1989, p. 261)) summarizes the three 'basic' problems for HMMs:

1. Given an observation sequence $O = O_1 O_2 \dots O_T$ and a model $\lambda = (A, B, \pi)$, how do we compute efficiently the probability of the observation sequence $P(O|\lambda)$?

2. Given an observation sequence $O = O_1 O_2 \dots O_T$ and the model λ, how do we choose a corresponding state sequence $Q = q_1 q_2 \dots q_T$ which is optimal in some meaningful sense (i.e. best "explains" the observations?

3. How do we adjust the model parameters $\lambda = (A, B, \pi)$ to maximise $P(O|\lambda)$?

Solving the first problem requires a forward-backward numerical procedure. The Viterbi algorithm (**?**) solves the second problem. In terms of structure, the Viterbi algorithm is similar to a forward procedure developed to solve the first problem, but it also includes a maximization at each node and a backtracking step.

The third problem is the most difficult of the three. The standard way of solving it is due to Baum and his coauthors and is known as the Baum-Welch algorithm (Baum et al., 1970, and references within). The Baum-Welch algorithm is in fact a special case of another celebrated algorithm: the EM or Expectation-Maximization algorithm (Dempster et al., 1977).

The Baum-Welch algorithm works by iteratively choosing a set of parameters $\lambda = (A, B, \pi)$ to maximise $P(O|\lambda)$. The iterative reestimation procedure is shown to converge monotonically to a <u>local</u> maximum. Like the EM algorithm, the Baum-Welch algorithm only identifies local maxima. From a practical perspective, this means that it is important to run the algorithm multiple times with different starting values to ensure that the solution obtained is the global maximum and not just a local one.

A.3 Model Selection

One of the difficulties with HMM models is to select the optimal number of states. We cannot use the Likelihood or the Loglikelihood directly, because the likelihood will increase as we increase the number of states. One way of addressing this problem is by selecting the model that optimzes one of the following information criteria:

1. The Akaika information criterion (AIC) is

$$AIC = -2 \ln L + 2p \tag{A.8}$$

   where L denotes the likelihood of the model and p is the number of parameters.

2. Schwartz' Bayesian information criterion (BIC) is

$$SBIC = -2 \ln L + 2p \ln T \tag{A.9}$$

   where T is the number of observations.

3. The Hannan-Quinn criterion (HQIC) is

$$HQIC = -2 \ln L + 2p \ln(\ln(T)) \tag{A.10}$$

   where T is the number of observations.

All three information criteria maximize the log likelihood of the model penalized by the number of parameters. The key difference between the three criteria relates to the treatment of the number of observations. While the Akaika information criterion ignores the number of observations, the Bayesian information criterion penalizes by the log of the number of observations. The Hannan-Quinn criterion is in between: the penalty is linked to the number of observations, but it is less stiff than the Bayesian information criterion.

In this paper, we consider both the AIC and BIC to determine the optimal number of states.

References

D. Abreu and M. K. Brunnermeier. Bubbles and crashes. *Econometrica*, 71 (1):173–204, 2003.

F. Allen and G. Gorton. Churning bubbles. *Review of Economic Studies*, 60 (4):813–836, 1993.

E. B. Andrade, T. Odean, and S. Lin. Bubbling with excitement: An experiment. *Review of Finance*, 20(2):447–466, 2016. doi: 10.1093/rof/rfv016. URL `http://rof.oxfordjournals.org/content/20/2/447.abstract`.

M. Baker and J. Wurgler. Investor sentiment and the cross-section of stock returns. *Journal of Finance*, 61(4):1645–1680, 2006.

A. Balkema and L. de Haan. Residual life time at great age. *Annals of Probability*, 2:792–804, 1974.

G. Barone-Adesi, L. Mancini, and H. Shefrin. A tale of two investors: Estimating optimism and overconfidence. Technical report, 2013.

L. Baum, T. Petrie, G. Soules, and N. Weiss. A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The Annals of Mathematical Statistics*, 41(1):164–171, 1970.

K. Berge, G. Consigli, and W. T. Ziemba. The predictive ability of the Bond-Stock Earnings Yield Differential. *Journal of Portfolio Management*, 34 (3):63–80, 2008.

O. J. Blanchard and M. W. Watson. Bubbles, rational expectations, and financial markets. In P. Wachtel, editor, *Crisis in the Economic and Financial Structure*, pages 295–315. 1982.

W. Buffett and C. Loomis. Mr. Buffett on the stock market. *FORTUNE Magazine*, November 22nd 1999.

W. Buffett and C. Loomis. Warren Buffett on the stock market. *FORTUNE Magazine*, December, 10 2001.

C. Camerer. Bubbles and fads in asset prices. *Journal of Economic Survey*, 3(1):3–14, 1989.

J. Y. Campbell and R. J. Shiller. Stock prices, earnings, and expected dividends. *The Journal of Finance*, 43(3):661–676, July 1988. Papers and Proceedings of the Forty-Seventh Annual Meeting of the American Finance Association, Chicago, Illinois, December 28-30, 1987.

S. Coles. *An Introduction to Statistical Modeling of Extreme Values*. Springer-Verlag, 2001.

R. Cont. Empirical properties of asset returns: stylized facts and statistical issues. *Quantitative Finance*, 1(2):223–236, 2001.

B. Corgnet, R. Hernán-González, P. Kujal, and D. Porter. The effect of earned versus house money on price bubble formation in experimental asset markets. *Review of Finance*, 19(4):1455–1488, 2015. doi: 10.1093/rof/rfu031. URL `http://rof.oxfordjournals.org/content/19/4/1455.abstract`.

A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum-likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Series B*, 39(1):1–38, 1977.

B. T. Diba and H. I. Grossman. The theory of rational bubbles in stock prices. *Economic Journal*, 98(392):746–754, 1988.

P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer-Verlag, London, 2011.

K. Fisher and M. Statman. Investor sentiment and stock returns. *Financial Analyst Journal*, 56(2):16–23, March/April 2000.

K. Fisher and M. Statman. Consumer confidence and stock returns. *Journal of Portfolio Management*, 30(1):115–127, Fall 2003.

R. A. Fisher. *Statistical Methods for Research Workers.* Oliver & Boyd, Edinburgh and London, 5 edition, 1933.

R. P. Flood, R. J. Hodrick, and P. Kaplan. An evaluation of recent evidence on stock market bubbles. NBER Working Paper No. w1971, 1986.

W. Goetzmann, D. Kim, and R. Shiller. Crash beliefs from investor surveys. 2016.

B. Graham and D. L. Dodd. *Security Analysis.* Whittlesey House, 1 edition, 1934.

F. Gresnigt, E. Kole, and P.H. Franses. Interpreting financial market crashes as earthquakes: A new early warning system for medium term crashes. *Journal of Banking & Finance*, 56(123–139), 2015.

G. Grimmett and D. Stirzaker. *Probability and Random Processes.* Oxford University Press, 2001.

R. A. Jarrow. Detecting asset price bubbles. *Journal of Derivatives*, 20(1): 30–34, 2012.

R. A. Jarrow, Y. Kchia, and P. Protter. A real time bubble detection methodology. *Bloomberg Risk Newsletter*, 2011a.

R. A. Jarrow, Y. Kchia, and P. Protter. How to detect an asset bubble. *SIAM Journal on Financial Mathematics*, 2:839–865, 2011b.

R. A. Jarrow, Y. Kchia, and P. Protter. Is there a bubble in Linkedin's stock price? *Journal of Portfolio Management*, 38(1):125–130, 2011c.

S. Lleo and W. T. Ziemba. Stock market crashes in 2007–2009: were we able to predict them? *Quantitative Finance*, 12(8):1161–1187, 2012.

S. Lleo and W. T. Ziemba. Some historical perspectives on the bond-stock yield model for crash prediction around the world. *International Journal of Forecasting*, 31(2):399–425, 2015.

S. Lleo and W. T. Ziemba. Does the bond-stock earnings yield differential model predict equity market corrections better than high P/E models? *Financial Markets, Institutions & Instruments*, 26(2):61–123, May 2017.

S. Lleo and W. T. Ziemba. Can Warren Buffett forecast equity market corrections? Technical report, 2018.

J. Neyman. On the two different aspects of the representative method: The method of stratified sampling and the method of purposive selection. *Journal of the Royal Statistical Society*, 97(4):558–625, 1934.

J. Neyman. Outline of a theory of statistical estimation based on the classical theory of probability. *Philosophical Transactions of the Royal Society A*, 236(767):333–380, 1937.

J. Neyman and E. S. Pearson. The testing of statistical hypotheses in relation to probabilities a priori. 29:492–510, 1933.

Y. Ogata. Statistical models for earthquake occurrences and residual analysis for point processes. *Journal of the American Statistical Association*, 83 (401):9–27, 1988.

J. Pickands. Statistical inference using extreme order statistics. *Annals of Statistics*, 3:119–131, 1975.

L. Rabiner. A tutorial in Hidden Markov Models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2):257–286, February 1989.

L. Rabiner and B. H. Juang. *Fundamentals of Speech Recognition*. Prentice Hall Signal Processing Series. Prentice Hall, 1993.

Y. Sato. Fund tournaments and asset bubbles. *Review of Finance*, 20(4):1383–1426, 2016. doi: 10.1093/rof/rfv044. URL `http://rof.oxfordjournals.org/content/20/4/1383.abstract`.

R. J. Shiller. *Irrational Exuberance*. Princeton University Press, 1 edition, 2005.

A. .N. Shiryaev and M. V. Zhitlukhin. Bayesian disorder detection problems on filtered probability spaces. *Theory of Probability and Its Applications*, 57(3):497–511, 2012a.

A. .N. Shiryaev and M. V. Zhitlukhin. Optimal stopping problems for a brownian motion with a disorder on a finite interval. Technical report, arXiv:1212.3709, 2012b.

A. .N. Shiryaev, M. V. Zhitlukhin, and W. T. Ziemba. When to sell Apple and the Nasdaq 100? trading bubbles with a stochastic disorder model. *Journal of Portfolio Management*, 40(2):1–10, Winter 2014.

A. .N. Shiryaev, M. V. Zhitlukhin, and W. T. Ziemba. Land and stock bubbles, crashes and exit strategies in japan, circa 1990 and in 2013. *Quantitative Finance*, 15(9):1449–1469, 2015.

W. T. Ziemba and S. L. Schwartz. *Invest Japan*. Probus, Chicago, 1991.

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

E·S·R·C ECONOMIC & SOCIAL RESEARCH COUNCIL

Systemic Risk Centre

The London School of Economics
and Political Science
Houghton Street
London WC2A 2AE
United Kingdom

Tel: +44 (0)20 7405 7686
systemicrisk.ac.uk
src@lse.ac.uk