Connected Stocks

Miguel Antón and Christopher Polk[1]

LONDON SCHOOL OF ECONOMICS

First draft: May 2008
This version: March 2010

[1] Antón: Department of Finance, London School of Economics, London WC2A 2AE, UK. Email m.anton1@lse.ac.uk. Polk: Department of Finance, London School of Economics, London WC2A 2AE, UK. Email c.polk@lse.ac.uk. We thank participants at the LSE lunchtime workshop, the Harvard PhD brownbag lunch, the HEC 2nd Annual Hedge Fund Conference, John Campbell, Randy Cohen, Owen Lamont, Augustin Landier, Jeremy Stein, and Tuomo Vuolteenaho for helpful comments. Financial support from the Paul Woolley Centre at the LSE is gratefully acknowledged. Antón also gratefully acknowledges support from the Fundación Ramón Areces.

Connected Stocks

Abstract

Recent theories suggest that institutional features may play an important role in the movement of stocks' discount rates, causing returns to comove above and beyond that implied by their fundamentals. We exploit the information in institutional connections to forecast cross-sectional variation in the extent to which stocks covary together. We connect stocks through common ownership by active mutual funds and find that common ownership by these funds predicts higher covariance, controlling for standard characteristics such as similarity along the dimensions of industry, size, book-to-market ratio, and momentum as well as the extent to which a pair of stocks are connected through common analyst coverage. The predictive effect is statistically and economically quite significant. We provide evidence that the comovement arising from ownership connections is due to a contagion effect. First, common ownership has a stronger effect on subsequent covariation when the stocks in the pair are small and/or the common owners are experiencing either strong inflows or outflows. Second, a decomposition of the covariation into cash-flow and discount-rate news components reveals that much of the aforementioned patterns are due to the interaction between the cash-flow news of one stock in the pair and the discount-rate news of the other stock in the pair. Finally, a trading strategy that uses the return on stock's connected portfolio as a confirming signal for a short-term, cross-stock reversal effect generates abnormal returns of over 7% per year, controlling for the own-stock reversal effect as well as other standard cross-sectional patterns in average returns. We show that both the typical hedge fund and the typical long/short hedge fund covaries negatively with this strategy, suggesting that the typical hedge fund may be part of the problem (creating the covariance) instead of part of the solution.

JEL classification: G12, G14, N22

1 Introduction

Barberis and Shleifer (2003) and Barberis, Shleifer and Wurgler (2005) have argued that institutional features may play an important role in the movement of stocks' discount rates, causing returns to comove above and beyond that implied by their fundamentals. In this paper we propose a new way to document that type of institutional comovement. Specifically, we forecast the off-diagonal elements of the firm-level covariance matrix using measures of institutional connectedness. By measuring institutional comovement in such a bottoms-up fashion, we can more precisely measure the covariation linked to institutional features. We focus on connecting stocks through active fund ownership as that institution not only may reflect existing patterns in covariation but also may layer on additional covariation. In particular, we study how common ownership of two stocks by an active fund manager can forecast the pair-wise covariation of those stocks, controlling for various other characteristics of the pair.

We find that active fund connectedness predicts higher covariance, controlling for similarity along the dimensions of industry, size, book-to-market ratio, and momentum as well as the extent to which a pair of stocks are connected through common analyst coverage. The predictive effect is both statistically and economically quite significant. We provide evidence consistent with common ownership causing the increased covariation associated with ownership. First, common ownership has a stronger effect on subsequent covariation when the stocks in the pair are small and/or the common owners are experiencing either strong inflows or outflows. Second, a decomposition of the covariation into cash-flow and discount-rate news components reveals that much of the aforementioned patterns are due to the interaction between the cash-flow news of one stock in the pair and the discount-rate news of the other stock in the pair. Interestingly, the ability of common analyst coverage to predict cross-sectional variation in comovement is primarily due to the covariance of cash-flow news with cash-flow news.

Previous and current research looks at related questions: Is there information in institutional holdings about future returns? Or more particularly, does variation in assets under management result in price pressure? Most of these studies are concerned with cross-sectional and time series predictability of abnormal returns. Any implications for comovement are secondary, if examined at all. We begin by measuring comovement and then we turn to the implications for predictability of returns at the end of the analysis. Specifically, we measure a stock's connected return and show that this connected return predicts cross-sectional variation in average returns. Specifically, we define the connected return for a particular stock as the return on a portfolio consisting of all the stocks in our sample which are connected to a particular stock through common ownership.

We document that trading strategies using the return on a stock's connected portfolio as a confirming signal for a short-term, cross-stock reversal effect generate significant abnormal returns of over 7% per year, controlling for market, size, value, mometum, and the own-stock, short-term reversal factors. This evidence we provide is again consistent with ownership-based connections causing the comovement.

Finally, we use our connected return strategy to explain hedge fund index returns in standard performance attribution regressions. We show that both the typical hedge fund and the typical long-short hedge fund load negatively on our trading strategy. In fact, the exposure of these value-weight hedge fund indexes are more negative than the sensitivity to our strategy of a value-weight portfolio of the active mutual funds in our sample. This suggests that hedge funds do not take advantage of the opportunities that arise from ownership-based connections. In fact, this finding suggests that the typical hedge fund may be part of the problem (creating the covariance) instead of part of the solution.[2]

Our work builds on a growing literature. It is now well known that there is a relation between mutual fund flows and past performance (Ippolito (1992), Chevalier and Ellison (1997), Sirri and Tufano (1998)). A recent paper by Coval and Stafford (2007) documents that extreme flows result in forced trading that temporarily moves prices away from fundamental value as in the general asset fire sales model of Shleifer and Vishney (1992) through the price pressure mechanism of Scholes (1972). Ellul, Jotikasthira, and Lundblad (2009) and Mitchell, Pedersen, and Pulvino (2007) document broadly similar findings in the bond and convertible bond markets respectively. Unlike these papers which study particular events, our analysis explores the extent to which institutional connections affect second moments more generally.

Recent theoretical work has emphasized the importance of delegated portfolio management and agency frictions to price movements such as these.[3] In particular,

[2] Consistent with this conclusion, Ben-David, Franzoni, and Moussawi (2009) argue that hedge funds consume rather than provide liquidity.

[3] See, for example, Darrell Duffie's 2010 AFA presidential address.

Vayanos and Woolley (2008) show how fund flows can generate comovement and lead-lag effects of the type we document. Their model provides strong theoretical motivation for our empirical analysis. More generally, beginning with Shleifer and Vishny (1997), researchers have studied the role of funding in arbitrage activity and the extent to which arbitrageurs should be expected to demand or provide liquidity.[4] On a related issue, Sadka (2009) shows that the typical hedge fund loads on a liquidity risk factor and that sensitivity to that liquidity risk is priced in the cross section of hedge fund returns. Measuring the extent to which hedge funds' performance can be attributed to a trading strategy that exploits temporary price dislocations due to institional-driven comovement follows naturally from that theory and empirical evidence

Four recent working papers analyze issues related to stock return comovement and/or insitutional ownership. Lou (2009) shows that flow-driven demand shocks more generally affect prices than just in the extreme fire-sale situations of Coval and Stafford and that in fact that mechanism goes a long way to explaining mutual fund performance persistance, the smart money effect, and price momentum among large-cap stocks. Unlike Lou (or Coval and Stafford for that matter), we avoid having to measuring the impact of flows on stock returns and instead use the actual connected return as a signal of the strength of the contagion effect resulting from ownership-based connections in the stock market. Moreover, whereas Lou's focus is on momentum effects, we instead examine how the presence of institutional connectedness interacts with the short-term reversal effect found in stock returns.

Sun (2008) uses standard clustering techniques to identify subsets of funds that hold similar stocks. Sun shows that the typical stock's return covarys with the equal-weight average return on all of the stocks in the top five fund clusters holding the stock in question. Moreover, Sun shows that this covariance is stronger if the average flow for the top five clusters in question is lower than the tenth percentile of the historical distibution of fund flows for that group of five fund clusters. In contrast, our approach models the pair-specific covariation as a function of the number of common funds holding the stock, controlling for style effects. Additionally, Sun does not examine any implications of the covariance she documents for profitable trading strategies.

Chen, Chen, and Li (2009) study the determinants of cross-sectional variation in

[4]Many researchers have built on the ideas in Shleifer and Vishny (1997), including Gromb and Vayanos (2002), Vayanos (2004), and Brunnermeier and Pedersen (2009).

pair-wise correlations and show that a large portion of that cross-sectional variation is persistant, yet unexplained by a long list of variables. They do not use the degree of active fund ownership to connect stocks. Like us, Chen, Chen, and Li develop a trading strategy that uses the return on the portfolio of stocks that comove with the stock in question. However, their trading strategy is a momentum strategy – buy (sell) stocks that have a high (low) comover's return. In contrast, our strategy is a contrarian one – sell (buy) stocks that have a high (low) connected portfolio return.

A paper written subsequent to our work that builds on our analysis is Greenwood and Thesmar (2009). Greenwood and Thesmar point out that owners of stocks can have correlated trading needs and thus the stocks that they hold can comove, even if there are no overlapping holdings. Greenwood and Thesmar show that these correlated trading needs predict future price volatility and cross-sectional variation in comovement.

Chen, Hanson, Hong, and Stein (2008) explores whether hedge funds take advantage of the mutual fund forced trading that Coval and Stafford document. They argue that hedge funds take advantage of that opportunity as average returns of long-short hedge funds are higher in months when the number of mutual funds in distress is large. In particular, Chen, Hanson, Hong , and Stein suggest that this evidence is consistent with hedge funds front-running the trades of distressed mutual funds. In contrast, our findings suggest that the typical hedge fund is apparently on the wrong side of the price dislocation that we study.

In summary, we show that understanding connectedness is a simple way to identify institutional-based stock comovement and its link to short-term reversal patterns. The rest of the paper is organized as follows. In Section 2, we summarize our methodology and data sources. In Section 3, we describe our results. Section 4 concludes.

2 Methodology

2.1 Measuring Commonality

We measure the amount of comovement in each pair that can be described by commonality in active mutual funds and equity analysts. At each quarter-end, we measure the number of funds ($F_{ij,t}$) that held both stocks i and j in their portfolios. As recent work by Brown, Wei, and Wermers (2009) suggests that analyst recommendations facilitate herding by mutual fund managers, we create similar measures of common analyst coverage. Specifically, we measure the number of analysts ($A_{ij,t}$) that issued at least one earnings forecast for both stocks i and j during the twelve month period preceding t. We use annual forecasts for our measure of common coverage as quarterly earnings forecasts are not issued as consistently. For each cross section, we calculate the normalized (to have unit standard deviation) rank transform of $F_{i,j}$ and $A_{i,j}$ which we denote as $F^*_{ij,t}$ and $A^*_{ij,t}$.

2.2 Modeling Cross-Sectional Variation in Comovement

To measure how commonality is linked to comovement, we estimate cross-sectional regressions forecasting subsequent cross-products of monthly returns for each pair of stocks. We initially forecast cross products of returns rather than cross products of unexpected returns because means are difficult to measure (Merton (1980)).

Our goal is to determine whether institutional connectedness contributes to a benchmark forecast of second moments. This is because one might expect that covariation, whether due to fundamentals or not, can be linked to the characteristics of two firms in a pair. The prototypical example is industry classification; we expect firms in similar industries to covary more, all else equal. To capture that similarity, we measure industry similarity as the number of consecutive SIC digits that are equal for a given pair.

In addition to industry similarity, we use three characteristics that help explain differences in the cross-section of returns, namely, size, book-to-market, and momentum. Previous research by Fama and French (1993) and Carhart (1997) has documented the link between these characteristics and common return factors. Therefore, we expect higher correlation between two stocks if they have a higher similarity in the characteristics mentioned above. To measure this similarity, we calculate percentile breakpoints for each characteristic and measure the difference in percentiles for a pair. As with our institutional measures, we use normalized rank transforms of $DIFF\_SIZE$, $DIFF\_BEME$, $DIFF\_MOM$, and NUM_SIC in our regressions. We denote those transformed variables with an asterisk superscript. For $DIFF\_SIZE$, $DIFF\_BEME$, and $DIFF\_MOM$, we employ the negative of the transformation in our regression analysis so that relatively high values of these measures indicate relatively more style similarity within the pair and therefore relatively higher comovement. As institutional ownership is correlated with size, we also create very general size controls based on the normalized rank transform of the percentile market capitalization of the two stocks, $SIZE1$ and $SIZE2$ (where we label the larger stock in the pair as the first stock), and the interaction between the two market capitalization percentile rankings.

The benchmark forecasting cross-sectional regression that we estimate is therefore the following:

$$\begin{aligned} r_{i,t+1}r_{j,t+1} &= a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} + b_s * DIFF\_SIZE^*_{ij,t} \\ &\quad + b_b * DIFF\_BEME^*_{ij,t} + b_m * DIFF\_MOM^*_{ij,t} + b_k * NUM\_SIC^*_{ij,t} \\ &\quad + b_{s1} * SIZE1_{ij,t} + b_{s2} * SIZE2_{ij,t} + b_{s12} * SIZE1SIZE2^*_{ij,t} + \varepsilon_{ij,t}. \end{aligned} \tag{1}$$

The dependent variable is the cross-product of returns at time $t+1$, updated monthly. The terms on the right hand side are measured at t and are all updated quarterly. We also estimate an alternative specification:

$$\begin{aligned} r_{i,t+1}r_{j,t+1} &= a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} \\ &\quad + \sum_{s=0}^{9} b_s * D_{DIFF\_SIZE_{ij,t}=s} + \sum_{b=0}^{9} b_b * D_{DIFF\_BEME_{ij,t}=b} \\ &\quad + \sum_{m=0}^{9} b_m * D_{DIFF\_MOM_{ij,t}=m} + \sum_{k=0}^{3} b_k * D_{NUM\_SIC_{ij,t}=k} \\ &\quad + b_{s1} * SIZE1_{ij,t} + b_{s2} * SIZE2_{ij,t} + b_{s12} * SIZE1SIZE2^*_{ij,t} + \varepsilon_{ij,t}. \end{aligned} \tag{2}$$

In this version of the regression, our control variables for a pair's difference in location across size, book-to-market, and momentum deciles as well as similarity in SIC code at the first, second, third, and fourth digit are allowed to come in through a simple but flexible dummy-variable specification.

In both cases, we estimate these coefficients using the approach of Fama and McBeth (1973). All independent variables are cross-sectionally demeaned so that the intercept a measures the average cross-sectional effect. We calculate Newey-West standard errors of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes out to four lags.

2.3 Data and Sample

Stock returns come from the monthly file in CRSP. We use common stocks (share codes 10 and 11) from NYSE, AMEX and NASDAQ whose market capitalization is above the NYSE median market cap. We choose this screening criteria because common ownership by active managers and common coverage by analysts is not pervasive: small stocks, especially in the beginning of the sample, have little institutional ownership in general. Limiting the data in this way also keeps the sample relatively homogeneous.

The data on mutual fund holdings come from the merge between CDA/Spectrum database provided by Thomson Reuters and CRSP Mutual Fund database. We use the Mutual Fund Links dataset created by Russ Wemers and offered by Wharton Research Data Services. As our focus is on US active mutual funds, we remove index, tax-managed funds and international funds by applying standard screening criteria used in the literature.[5] In addition, for a fund to be in our sample we require it to hold at least one stock of our stock sample.

We obtain data on analysts from the Institutional Brokers Estimate System (I/B/E/S) database. At each point in time, we observe the stocks covered by each analyst through the earnings forecasts that they issue. For an analyst to be in our sample, we require that he or she follow at least one of the stocks in our stock sample by issuing a one-year earnings forecast (the most common forecast issued by an analyst).

Our sample covers the period 1983 to 2007. Table 1 confirms the well-known marked increase in funds over this period. The number of analysts has also increased, though not as dramatically. Tables 2 and 3 report estimates of aggregate and firm-level VARs. These estimates allow us to decompose returns into their cash-flow news and discount-rate news components using the approach of Campbell (1991). We summarize his method and the particular VAR specifications that we use to implement his technique in the Appendix. Table 4 reports various summary statistics for returns and the news components. Consistent with Vuolteenaho (2002), cash-flow

[5] We specifically follow the algorithm described in Cremers and Petajisto (2009).

news makes up a larger portion of total return variance.

3 Results

Table 4 measures the extent of active managers' and analysts' workloads. For these active managers, the median load is 40 above-median NYSE capitalization stocks. For analysts, the median load over this subset of stocks is five firms. Consequently, this workload results in typically 16 analysts covering a firm and 37 funds holding the stock of that firm. Because of the growth of funds over this period, these full-sample numbers mask a strong trend in the number of funds holding a stock. In the early part of the sample (1983-1989), the median number of funds holding one of the above-median NYSE capitalization stocks was 9. In the later part of the sample (2000-2007), that median number increased to 102.

Our specific interest is how these numbers translate to the number of common owners or the amount of common coverage for a pair of stocks. We report those numbers in Tables 5 and 6. In terms of coverage, it is quite rare to share an analyst with another firm. In fact, only 5% of all pairs have an analyst in common. In contrast, it is relatively common to share active fund ownership with another stock as more than 75% of all stock pairs have a common active fund owner. Typically a pair would have roughly seven funds in common. Table 5 shows that the number of ownership-based connections among above-median NYSE capitalization stocks has increased dramatically over the period we study. In 1988, the median number of ownership connections was 3. In 2007, the median number of ownership connections was 19. Our use of only rank-transformed variables in the analysis is exactly because of this trend. Figure 1 plots how the average number of common owners in the cross section of pairs we study has evolved over time. For interpretability, we scale this measure by the expected number of common owners per pair under the assumptions that all funds hold the same number of stocks in our sample at a particular point in time as the average fund at that time. One can see that relative to this benchmark, the average number of connections has varied through time but has trended up over the sample period.

Table 7 reports the result of our forecasting cross-sectional variation in realized cross products. We begin by estimating simpler versions of equation (1). In column (1), we estimate a specification with only common ownership as a forecasting variable.

That variable is highly statistically significant, with a coefficient of 0.00030 and a t-statistic of 6.11. Recall that the common fund variable has been normalized to have a standard deviation of one and a mean of zero. Therefore the constant term, 0.00216, reflects the average realized cross product and is a useful benchmark to understand the economic significance of our finding. Specifically, the coefficient on common funds indicates that a change of one standard deviation in the degree of common ownership results in an increase in the forecasted cross product that is approximately 14.4% of the average amount of covariation. In column (2) of Table 7 , we predict covariation using our measure of common ownership and common coverage, absent any other controls. The coefficient on our measure of common funds is 0.00027 with a t-statistic of 5.73, only 10% smaller than the estimate in column (1). Thus there seems to be little correlation in the extent to which $F^*_{ij,t}$ and $A^*_{ij,t}$ drive cross-sectional variation in comovement. The coefficient on common analyst coverage, 0.00018, indicates that a one standard deviation increase in the amount of common analysts results in an increase in comovement of more than 8% of the average realized covariation. The coverage-based coefficient is also measured quite precisely with a t-statistic of 7.49.

Being able to forecast differences in comovement using institutional connectedness may not be surprising if the predictability simply reflects the fact that fund managers and analysts choose to hold stocks that are similar and therefore would be expected to comove regardless of the common ownership or coverage. For example, growth managers will tend to hold growth stocks, and previous research has shown that those types of stocks tend to covary. Therefore, we include four controls for whether the stocks in the pair are similar. Column (3) of Table 7 reports the result of that analysis. Recall that these control variables are normalized to have a standard deviation of one and transformed (in the case of size, book-to-market, and momentum) so that higher values indicate greater style similarity. We find a strong effect for a one-standard deviation move in industry similarity as the coefficient is 0.00020 with a t-statistic of 7.30. There is a relatively strong pattern for similarity in book-to-market as well. The coefficient associated with a one-standard deviation move in style similarity is 0.00012 (t-statistic of 2.78). The similarity in momentum has the same one-standard deviation effect on differences in comovement as the similarity in book-to-market (coefficient of 0.00012) with a slightly lower t-statistic of 2.28. The effect on comovement due to size is statistically indistinguishable from zero. More importantly, the coefficient on common ownership barely changes (0.00024, a drop of only 0.00003) and remains quite statistically significant. Interestingly, the coefficient on common ownership has the strongest one-standard-deviation influence among the variables under consideration.

In column (4), we estimate the full benchmark specification. Here we now include very general controls for the size of the stocks in the pair. All else equal, one might expect that having large stocks in the pair would increase comovement as these stocks will reflect more of the market's movements. More generally, one might think that size is very important in determining institutional ownership. Though these controls are important in describing cross-sectional variation in comovement, the institutional connectedness variables are still quite significant and in fact the measured coefficients become stronger.

The final column of Table 7 generalizes our controls for stock similarity by turning to dummy variables to capture the difference in size, beme, or momentum decile across the pair. We also dummy the number of common SIC digits. We report these dummy coefficient estimates of equation (2) in Panel B of Table 7. The results show that this flexibility appears to be important. For example, the increase in comovement when a pair goes from having zero to one SIC digit in common is much more important than going from having two to three SIC digits in common. Nevertheless, this more flexible specification does not affect the coefficient on our common ownership variable.

In Table 8, we use alternative measures of comovement between two stocks. In the first column of Table 8, we repeat the estimates from the fourth column of Table 7 (our full specification) for ease of comparison. In column (2), we keep the same control variables as in the full benchmark specification of Table 7 but replace the monthly return cross product with the corrected sum of daily return cross products ($S_{r_i r_j} = \sum_{k=1}^{N} r_{i,k} r_{j,k} - \frac{1}{N} \sum_{k=1}^{N} r_{i,k} \sum_{k=1}^{N} r_{j,k}$) for the N days within month t+1. We find that the coefficient on $F^*_{ij,t}$ has much more statistical significance (t-statistic of 9.05) and continues to be quite economically significant (20% of the average effect, as estimated by the constant term). The increase in statistical significance is consistent with the notion that high-frequency estimates of second moments are more precise. In columns (3) and (4), we again keep the same control variables as in Table 7 but replace the monthly return cross product with Fisher and Pearson measures of the correlation coefficient of the daily returns on stock i and j within month t+1. The coefficient remains economically large and has a t-statistic over 16 in both cases. This result confirms that our measure of connectedness forecasts cross-sectional variation in correlation. Taken together, the results in Table 8 ease concerns of our use of the realized monthly return cross product (and its components) throughout the rest of the paper.

To summarize, the main conclusion from Table 7 is that institutional connectedness, whether through common coverage or common ownership, gives economically and statistically significant ability to forecast subsequent comovement. It is worth noticing that we are only examining in-sample forecasting of cross-sectional variation in the covariance matrix. However, given that the literature currently concludes that 1/N rules are about the best one can do out-of-sample, it would be interesting to explore how our method and our characteristics perform in out-of-sample tests such as those in DeMiguel, Garlappi, and Uppal (2007). Since the characteristics we are using are relatively persistent, we expect that our method and model will perform relatively well out-of-sample, consistent with the related claims of Brandt, Santa-Clara, and Valkanov (2009).

3.1 When does connectedness matter?

Table 7 documents that institutional connectedness helps predict cross-sectional variation in comovement. The rest of the analysis will focus on exploring why connecting stocks through common fund ownership matters. A likely explanation is that the effect we find is consistent with a causal relationship due to price pressure arising from flows (Coval and Stafford (2007), Lou (2009)). To provide additional evidence that this is the case, we exploit cross-sectional heterogeneity in stock pair characteristics. Specifically, in Table 9 we interact the coefficient on common funds with dummies for the size of the pair of stocks and the total net flow into the common funds. Specifically, each quarter we sort pairs into quintiles based on their total market capitalization. We independently sort pairs into quintiles based on their total net flow. We follow the literature in defining flows (see Coval and Stafford, 2007). Therefore, the net relative investment flow of funds into fund i in quarter t is defined as:

$$FLOW_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} * (1 + R_{i,t})}{TNA_{i,t-1}} \tag{3}$$

where $TNA_{i,t}$ is the Total Net Assets of fund i in quarter t and $R_{i,t}$ is the fund return over the period $t-1$ to t reported by CRSP Mutual Fund Database. Fund flows are reported quarterly before 1991 and monthly thereafter. To compute the quarterly flows, we first compute the absolute monthly flows, then we sum them up and finally we divide them by the previous quarter TNA.

We find that common ownership effect on comovement is stronger for pairs of smaller stocks. In every row, there is a strong decline in the coefficient as we move to pairs of larger stocks. Moreover, we find that the common ownership effect on comovement is strong for low net flows and high net flows. The lowest estimate in each column always occurs in the fourth row. We generally find a stronger effect for inflows than for outflows, though for the largest pairs, this difference is not statistically significant. Figure 2 shows these patterns graphically.

3.2 Connectedness and temporary components of returns

To better understand how institutional connectedness may cause comovement, we decompose unexpected returns into discount-rate news and cash-flow news. Two firms can be correlated because shocks to their cash-flows move together, because shocks to their discount rates move together, or because the shocks to the cash-flows of one firm moves with the shocks to the discount-rates of the other firm. What is useful about this decomposition in this context is that institutions cannot directly affect fundamentals. Therefore, predicting this portion of the decomposition clearly reflects the endogenous choice of institutions. Of course, a higher return covariance arising from higher covariance between the discount-rate news of the pair is also consistent with plausible endogeneity-based explanations. For example, firms may tend to hold pairs that load on a particular priced common factor, not captured by size, book-to-market, or momentum, whose expected return varies through time. Consider, however, covariation between the cash-flow news of one firm and the discount-rate news of another. This covariation predictability seems much more difficult to explain away as simply reflecting the endogenous choice of the fund manager and seems quite more likely to be due to institutions having a causal role. The likely mechanism, given the results in Table 9, would appear to be the price pressure story behind the results of Coval and Stafford (2007).

The methodology we now follow is very similar to the one described above, but we change the left hand side of equation 1. Specifically, the new equation we estimate has the form:

$$
\begin{aligned}
y_{ij,t+1} &= \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (4)\\
&\quad [N_{i,CF_{t+1}}N_{j,CF_{t+1}}; -N_{i,CF_{t+1}}N_{j,DR_{t+1}} - N_{j,CF_{t+1}}N_{i,DR_{t+1}}; N_{i,DR_{t+1}}N_{j,DR_{t+1}}] \\
&= a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} + b_s * DIFF\_SIZE^*_{ij,t} \\
&\quad + b_b * DIFF\_BEME^*_{ij,t} + b_m * DIFF\_MOM^*_{ij,t} + b_k * NUM\_SIC^*_{ij,t} \\
&\quad + b_{s1} * SIZE1_{ij,t} + b_{s2} * SIZE2_{ij,t} + b_{s12} * SIZE1 * SIZE2^*_{ij,t} + \varepsilon_{ij,t}
\end{aligned}
$$

where $y_{ij,t+1}$ is a vector of the various components of the realized return cross-product. The results of our covariance decomposition can be found in Table 10. We find that a significant proportion of the effect is due to the covariance of cash-flow news with cash-flow news. For the ownership-based connection, the estimate is a statistically significant 0.00018. As argued above, this component must reflect the choices that fund owners make. There is a statistically significant but economically small relation between common fund ownership and subsequent covariance between the discount-rate news of one stock in the pair and the discount-rate news of the other stock in the pair. However, consistent with the price pressure explanation, common fund ownership has a statistically significant relation with the covariation between cash-flow news and discount-rate news of the stocks in the pair. The measured coefficient is 0.00017, with a t-statistic of 4.27. Note that the average effect is -.00112. Thus for the typical stock pair, the interaction between cash-flow news for one stock and discount-rate news for the other stock tends to reduce return covariance between the stocks in the pair, but for stocks with common coverage, return covariation is increased.

In Table 11, we repeat our exercise of interacting the coefficient on $F^*_{ij,t}$ with dummies for the pairs location in sorts based on the size of the pair of stocks and the total net flow into the common funds. Consistent with our interpretation, Panel B of Table 11 shows that the cross-sectional variation in the magnitude of the coefficient documented in Table 9 also shows up in the $-N_{i,CF_{t+1}} * N_{j,DR_{t+1}}$ component of the return covariance, though somewhat less distinct than before.

3.3 Robustness to additional controls and measures of common ownership

Our regressions have controlled for similarity in characteristics that are known to describe variation in fund managers' investing themes. A recent paper by Chen, Chen, and Li (2009) documents that variables other than similarity in these characteristics forecast cross-sectional variation in pair-wise correlations. As a further robustness test, we control for their long list of pair characteristics. In particular, we include past return correlation, $RETCORR_{ij,t}$; past profitability correlation, $ROECORR_{ij,t}$; the past correlation in the stocks abnormal trading volume, $VOLCORR_{ij,t}$; the absolute value of the difference in five-year log sales growth rates, $DIFFGROWTH_{ij,t}$; the absolute difference in financial leverage ratios (defined as long-term debt / total assets), $DIFFLEV_{ij,t}$; a dummy variable in the two firms are located in the same state; the absolute value of the difference in the two stocks' log share prices, $DIFFPRICE_{ij,t}$; $D_{STATE_{ij,t}}$; a dummy variable if the two stocks belong to the same S&P major, mid-cap, or small-cap index, $D_{INDEXij,t}$; and a dummy variable if the two stocks are on the same stock exchange, $D_{LISTING_{ij,t}}$.

Table 12 repeats the key regressions from Tables 7, 8, and 10 but including these additional controls. In particular, in regression 2 of Table 12, we reproduce the main findings of Chen, Chen, and Li (2009). Stock pairs with relatively higher past return, profitability, or volume correlation have relatively higher return correlation in the future. Stock pairs that are located in the same state and belong to the same S&P index have relatively higher return correlation. (In contrast to Chen, Chen, and Li, though we find that stocks that trade on the same exchange do tend to have higher return correlation in the future, that effect is not statistically significant.) Stock pairs that are relatively more similar in their past sales growth rates, their current share price, or their current leverage ratio have relatively higher correlation in the future. None of these empirical regularities subsume our finding that two stocks with relatively higher common ownership have predictably higher subsequent comovement. Moreover, this higher comovement can be primarily attributed to the interaction between cash-flow news for one stock and discount-rate news for the other stock.

Table 13 adds these additional controls to the key regressions from Tables 9 and 11 that interact the coefficient on $F^*_{ij,t}$ with dummies for the pair's location in sorts based on the size of the pair of stocks and the total net flow into the common funds.

Panel A of Table 13 forecasts cross-sectional variation in subsequent return cross products. Panel B of Table 13 forecasts cross-sectional variation in the component of subsequent return cross products that is related to the covariation between cash-flow news and discount-rate news of the stocks in the pair. In both cases, we continue to find that the common ownership effect on comovement is stronger for extreme net flows as well as stronger for pairs of smaller stocks.

Table 14 varies the definition of common ownership for our benchmark specification (Panel A) and our specification that includes the Chen, Chen, and Li variables (Panel B). We first replace the number of common owners, $F_{ij,t}$, with the total ownership by all common funds in dollars of the two stocks scaled by the total market capitalization of the two stocks, $F_{ij,t}^{CAP}$. Our next alternative is to measure common ownership as the average ownership stake by all common owners across the two stocks, $F_{ij,t}^{SHARES}$. Finally, we turn to total ownership by all common funds in dollars of the two stocks scaled by the Total Net Assets of all common owners, $F_{ij,t}^{TNA}$, as our last measure. All definitions continue to forecast cross-sectional variation in the realized return cross-product, the subsequent return correlation, and the covariance between the cash-flow and discount-rate news components. Though differences in the relative forecasting ability appear relatively minor, it is comforting to see that our primary definition consistently has the largest t-statistic and provides the largest R^2.

3.4 Connected trading strategies

Here we measure the profits to a simple trading strategy based on our finding that ownership-based connectedness can be linked to temporary components of returns. If stock i experiences a negative cash flow shock and connected stock j's price also drops, we conjecture that the drop is due to price pressure, which we expect to revert. Our trading strategy is thus very simple: we buy (sell) stocks that have gone down (up) if their connected stocks have gone down (up) as well.

Each month, we sort our subset of stocks into quintiles based on past one-month return. We independently sort stocks into quitiles based on the past one-month return, $r_{iC,t}$, on their portfolio of connected stocks. We use $F_{ij,t}^{*}$ to generate the weights on the connected stocks in the portfolio. Define

$$\begin{aligned} F_{ij,t}^{**} &= F_{ij,t}^{*} \text{ if } F_{ij,t} > 0 \\ F_{ij,t}^{**} &= 0 \text{ if } F_{ij,t} = 0 \end{aligned}$$

Thus the return on the portfolio is $r_{iC,t} = \dfrac{\sum_{j=1}^{J} F^{**}_{ij,t-1} r_{j,t}}{\sum_{j=1}^{J} F^{**}_{ij,t-1}}$.

We consider two different trading strategies. The first strategy buys stocks that are in the low own-return and low connected-return portfolio while selling stocks that are in the high own-return and high connected-return portfolio. This strategy uses the connected return as a confirming signal of whether the own stock is under or overvalued. The second strategy buys stocks that are in the low own-return and high connected-return portfolio while selling stocks that are in the high own-return and low connected-return portfolio. Thus, the second strategy uses the connected return as a contrarian signal. For each strategy we generate the cumulative buy-and-hold abnormal return by regressing the $t+1, t+2, ..., t+12$ returns on the five-factor model

$$r_{p,t+1} - r_{f,t+1} = \alpha_5 + bRMRF_{t+1} + sSMB_{t+1} + hHML_{t+1} + mMOM_{t+1} + rSTREV_{t+1} + \varepsilon_{p,t+1}$$

where the factors are the four factors of Carhart (1997), augmented with the short-term reversal factor.[6] We include this factor as we are sorting the target stock on its past month return, though we also show results excluding that factor from our regression.

Figure 3 graphs the cumulative abnormal returns on these two different trading strategies. There are two important features of the graph. One, the average abnormal return in the first month after the sort is significantly higher when the connected return is used as a contrarian signal. Two, the cumulative average abnormal return is twice as large eight months after the sort when the connected stock return is used as a confirming signal. These two features are consistent with stocks being pushed away from fundamental value by mutual-fund trading, with the connected return being a useful measure of the extent of that temporary misvaluation. Thus, compared to the standard short-term reversal effect, the misvaluation is larger but takes more time to revert. Figure 4 emphasizes this difference. The trading strategies are the same as in Figure 3, except that we use the previous three-month return on a stock

[6] All factors are from Ken French's website.

and the previous three-month return on the connected portfolio. The cumulative buy-and-hold return when the connected return is used as a confirming signal rather than a contrarian signal is now nearly twice as large.

As a consequence, we design a trading strategy that takes these features into account. Table 15 reports the four and five-factor alphas from a trading strategy that uses the past three-month return on the own stock and its connected portfolio as a confirming signal and holds the stocks for five months. There are several general patterns that are consistent with the results in Figures 3 and 4. Holding the own return constant, as one moves from high to low connected return, alphas generally increase. Holding the connected return constant, as one moves from high to low own return, the alphas increase. The five-factor alpha of a strategy that buys the low own return and low connected return portfolio and sells the high own return and high connected return strategy earns an impressive 62 basis points per month. The corresponding t-statistic is 3.56. In Table 16, we also control for the liquidity factors of Sadka (2006) and Pastor and Stambaugh (2003), a linear time trend, and end-of-quater dummies. Our results are remain economically and statistically significant and are qualitatively the same.

3.5 Hedge Fund Index attribution

Our last analysis uses the connected stock trading strategy in performance attribution of hedge fund index returns using the CSFB/Tremont Hedge Fund Indexes. These indexes have been used in a number of studies including Asness, Krail, and Liew (2001); Agarwal and Naik (2004); Getmansky, Lo, and Makarov (2003); and Bondarenko (2004). We focus on two particular indexes. The first index is the index of all hedge funds. As CFSB weights hedge fund returns by assets under management and captures more than 85% of all assets under management, this index gives a good representation of the extent to which our connected stock strategy reflects the general health of the hedge fund industry.[7] We also examine the performance of the long/short component of the CSFB index to measure the extent to which funds which specifically invest in equities are exposed to the connected stocks factor.

Table 17 reports the results of this analysis. We find that hedge funds in general and long/short managers in particular load negatively on the connected stocks trading

[7] Note that the CFSB does not include managed accounts or funds of funds in its indexes.

strategy. The coefficient in the first column of Table 17 estimates a regression of the overall hedge fund index excess return on our connected return strategy and the four factors of Carhart (1997), augmented with the short-term reversal factor. The coefficient is -0.1006 with a t-statistic of -2.71. The second column of the Table instead attributes the performance of the hedge fund index to the connected strategy and the eight hedge fund factors of Fung and Hsieh (2004). Though hedge funds in the aggregate load on these eight factors to various degrees, our connected stocks factor remains important in describing the returns on hedge funds. The coefficient becomes both economically and statistically more significant as the point estimate drops to -0.1336 and the t-statistic to -6.33. This result suggests that our trading strategy is useful tool to measure the state of the hedge fund industry.

Column 3 of Table 17 measures the degree to which the Long/Short subset of hedge funds covary with our connected return trading strategy in the presence of the Carhart factors and the short-term reversal factor. We find a very similar exposure on the connected strategy as in column 1. When we instead use the Fung and Hsieh factors, we find that the returns on these funds strongly negatively covary with our connected return factor with loadings that are approximately 50% larger in absolute value. T-statistics are correspondingly bigger. For comparison, we also estimate the loading of a value-weight portfolio consisting of all of the active mutual funds in our sample over the same time period. This portfolio has a smaller (in absolute value) sensitivity to the connected strategy as the estimate is -0.0213 with an associated t-statistic of -1.90. Though we do not observe holdings data for hedge funds and therefore cannot see the exact positions of these long/short hedge funds, these results suggests that these hedge funds do not take advantage of the opportunities that price pressure from mutual fund flows provide. In fact, one can argue that perhaps hedge funds are excerbating rather than mitigating the price pressure patterns documented in this paper.

4 Conclusion

We show that stocks are connected through their common fund ownership. In particular, pairs of stocks that are connected in this fashion covary more together, controlling for similarity in industry, size, book-to-market equity ratio, and past return momentum as well as common analyst coverage. We present additional evidence that suggests the incremental comovement may be causal. First, the effect is stronger for pairs of relatively smaller stocks and is stronger for pairs whose common owners are experiencing strong inflows or outflows. Moreover, the effect flows through the interaction of cash-flow news for one stock with the discount-rate news of the other. Finally, trading strategies that exploit the fact that temporary price pressure must eventually revert are quite profitable. A trading strategy that uses the return on the portfolio of stocks that a particular stock is connected to as a confirming signal generates annual abnormal returns of over 7%. As a consequence, we provide a simple way to document the extent to which ownership-based connections result in equity market contagion. In an application, we document that hedge funds in general and an equity-focused subset in particular covary negatively with our trading strategy (and more so than the mutual funds we originally study), suggesting that hedge funds on average do not take advantage of the opportunities that contagion from ownership-based connections provides.

5 Appendix

5.1 Decomposing Stock Returns

The price of any asset can be written as a sum of its expected future cash flows, discounted to the present using a set of discount rates. Campbell and Shiller (1988a, 1988b) develop a loglinear approximate present-value relation that allows for time-varying discount rates. Campbell (1991) extends the loglinear present-value approach to obtain a decomposition of returns:

$$\begin{aligned} r_{t+1} - \mathrm{E}_t\, r_{t+1} &= (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \qquad (5) \\ &= N_{CF,t+1} - N_{DR,t+1}, \end{aligned}$$

where Δd denotes log dividend growth, r denotes log returns, N_{CF} denotes news about future cash flows (future dividends), and N_{DR} denotes news about future discount rates (i.e., expected returns). This equation says that unexpected stock returns must be associated with changes in expectations of future cash flows or discount rates.

5.2 Measuring the components of returns

An important issue is how to measure the shocks to cash flows and to discount rates. One approach, introduced by Campbell (1991), is to estimate the cash-flow-news and discount-rate-news series using a vector autoregressive (VAR) model. This VAR methodology first estimates the terms $\mathrm{E}_t\, r_{t+1}$ and $(\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}$ and then uses realization of r_{t+1} and equation (5) to back out cash-flow news. Because of the approximate identity linking returns, dividends, and stock prices, this approach yields results that are almost identical to those that are obtained by forecasting cash flows explicitly using the same information set. Thus the choice of variables to enter the VAR is the important decision in implementing this methodology.

When extracting the news terms in our empirical tests, we assume that the data are generated by a first-order VAR model

$$z_{t+1} = a + \Gamma z_t + u_{t+1}, \tag{6}$$

where z_{t+1} is a m-by-1 state vector with r_{t+1} as its first element, a and Γ are m-by-1 vector and m-by-m matrix of constant parameters, and u_{t+1} an i.i.d. m-by-1 vector of shocks.

Provided that the process in equation (6) generates the data, $t+1$ cash-flow and discount-rate news are linear functions of the $t+1$ shock vector:

$$\begin{aligned} N_{DR,t+1} &= e1'\lambda u_{t+1}, \\ N_{CF,t+1} &= (e1' + e1'\lambda)\, u_{t+1}. \end{aligned} \tag{7}$$

where $e1$ is a vector with first element equal to unity and the remaining elements equal to zeros. The VAR shocks are mapped to news by λ, defined as $\lambda \equiv \rho\Gamma(I - \rho\Gamma)^{-1}$ so that $e1'\lambda$ measures the long-run significance of each individual VAR shock to discount-rate expectations.

5.3 Aggregate VAR Specification

In specifying the monthly aggregate VAR, we follow Campbell and Vuolteenaho (2004), choosing the same four state variables that they study. The first element of our state vector is the excess log return on the market (r_M^e), the difference between the annual log return on the CRSP value-weighted stock index (r_M) and the annual log riskfree rate, obtained from Professor Kenneth French's website. The second element of our state vector is the term yield spread (TY), provided by Global Financial Data and computed as the yield difference between ten-year constant-maturity taxable bonds and short-term taxable notes, in percentage points. The third variable is the log smoothed price-earnings ratio (PE), the log of the price of the S&P 500 index divided by a ten-year trailing moving average of aggregate earnings of companies in the index. As in Campbell and Vuolteenaho (2004), we carefully remove the interpolation inherent in Shiller's construction of the variable to ensure the variable does not suffer from look-ahead bias. Our final variable is the small-stock value spread (VS), which we construct using the data made available by Professor Kenneth French on his web site. The portfolios, which are constructed at the end of each June, are the intersections of two portfolios formed on size (market equity, ME) and three portfolios formed on the ratio of book equity to market equity (BE/ME). We generate intermediate values of VS by accumulating total returns on the portfolios in question.

Table 2 reports the VAR model parameters, estimated using OLS. Each row of the table corresponds to a different equation of the VAR. The first five columns report coefficients on the five explanatory variables: a constant, and lags of the excess market return, term yield spread, price-earnings ratio, and small-stock value spread. OLS standard errors are reported in parentheses below the coefficients. The first row of Table 2 shows that all four of our VAR state variables have some ability to predict monthly excess returns on the aggregate stock market. In our sample, monthly market returns display momentum; the coefficient on the lagged excess market return is a statistically significant 0.1118 with a t-statistic of 3.52. The regression coefficient on past values of the term yield spread is positive, consistent with the findings of Keim and Stambaugh (1986), Campbell (1987), and Fama and French (1989), but with a t-statistic of only 1.6. The smoothed price-earnings ratio negatively predicts the return with a t-statistic of 3.42, consistent with the finding that various scaled-price variables forecast aggregate returns (Campbell and Shiller, 1988ab, 2003; Rozeff 1984; Fama and French 1988, 1989). Finally, the small-stock value spread negatively predicts the return with a t-statistic of 2.16, consistent with Brennan, Wang, and Xia (2001), Eleswarapu and Reinganum (2004), and Campbell and Vuolteenaho (2004). In summary, the estimated coefficients, both in terms of signs and t-statistics, are consistent with previous research.

The remaining rows of Table 2 summarize the dynamics of the explanatory variables. The term spread can be predicted with its own lagged value and the lagged small-stock value spread. The price-earnings ratio is highly persistent, with past returns adding some forecasting power. Finally, the small-stock value spread is highly persistent and approximately an AR(1) process.

5.4 Firm-level VAR Specification

We implement the main specification of our monthly firm-level VAR with the following three state variables. First, the log firm-level return (r_i) is the monthly log value-weight return on a firm's common stock equity. Following Vuolteenaho (2002), to avoid possible complications with the use of the log transformation, we unlever the stock by 10 percent; that is, we define the stock return as a portfolio consisting of 90 percent of the firm's common stock and a 10 percent investment in Treasury Bills. Our second state variable is the momentum of the stock (MOM), which we measure following Carhart (1997) as the cumulative return over the months $t-11$ to $t-1$. Our final firm-level state variable is the log book-to-market equity ratio (we denote the transformed quantity by BM in contrast to simple book-to-market that is denoted by BE/ME) as of the end of each month t.

We measure BE for the fiscal year ending in calendar year $t-1$, and ME (market value of equity) at the end of May of year t.[8] We update BE/ME over the subsequent eleven months by dividing by the cumulative gross return from the end of May to the month in question. We require each firm-year observation to have a valid past BE/ME ratio that must be positive in value. Moreover, in order to eliminate likely data errors, we censor the BE/ME variables of these firms to the range (.01,100) by adjusting the book value. To avoid influential observations created by the log transform, we first shrink the BE/ME towards one by defining $BM \equiv \log[(.9BE + .1ME)/ME]$.

[8] Following Fama and French, we define BE as stockholders' equity, plus balance sheet deferred taxes (COMPUSTAT data item 74) and investment tax credit (data item 208) (if available), plus post-retirement benefit liabilities (data item 330) (if available), minus the book value of preferred stock. Depending on availability, we use redemption (data item 56), liquidation (data item 10), or par value (data item 130) (in that order) for the book value of preferred stock. We calculate stockholders' equity used in the above formula as follows. We prefer the stockholders' equity number reported by Moody's, or COMPUSTAT (data item 216). If neither one is available, we measure stockholders' equity as the book value of common equity (data item 60), plus the book value of preferred stock. (Note that the preferred stock is added at this stage, because it is later subtracted in the book equity formula.) If common equity is not available, we compute stockholders' equity as the book value of assets (data item 6) minus total liabilities (data item 181), all from COMPUSTAT.

The firm-level VAR generates market-adjusted cash-flow and discount-rate news for each firm each month. We remove month-specific means from the state variables by subtracting $r_{M,t}$ from $r_{i,t}$ and cross-sectional means from $MOM_{i,t}$ and $BM_{i,t}$. As in Campbell, Polk, and Vuolteenaho (2010), instead of subtracting the equal-weight cross-sectional mean from $r_{i,t}$, we subtract the log value-weight CRSP index return instead, because this will allow us to undo the market adjustment simply by adding back the cash-flow and discount-rate news extracted from the aggregate VAR.

After cross-sectionally demeaning the data, we estimate the coefficients of the firm-level VAR using WLS. Specifically, we multiply each observation by the inverse of the number of cross-sectional observation that year, thus weighting each cross-section equally. This ensures that our estimates are not dominated by the large cross sections near the end of the sample period. We impose zero intercepts on all state variables, even though the market-adjusted returns do not necessarily have a zero mean in each sample. Allowing for a free intercept does not alter any of our results in a measurable way.

Parameter estimates, presented in Table 3, imply that expected returns are high when past one-month return is low and when the book-to-market ratio and momentum are high. Book-to-market is the statistically most significant predictor, while the firm's own stock return is the statistically least significant predictor. Momentum is high when past stock return and past momentum are high and the book-to-market ratio is low. The book-to-market ratio is quite persistent. Controlling for past book-to-market, expected future book-to-market ratio is high when the past monthly return is high and past momentum is low.

References

Barberis, Nicholas and Andrei Shleifer, 2003, Style investing, *Journal of Financial Economics* 68, 161–199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler, 2005, Comovement, *Journal of Financial Economics* 75, 283–317.

Ben-David, Itzhak, Francesco Franzoni, and Rabih Moussawi, 2009, Do hedge funds provide liquidity?, Ohio State working paper.

Brandt, Michael W., Pedro Santa-Clara, and Rossen Valkanov, 2009, Parametric Portfolio Policies: Exploiting Characteristics in the Cross-Section of Equity Returns, *Review of Financial Studies*, forthcoming.

Brennan, Michael J., Ashley Wang, and Yihong Xia, 2001, A simple model of intertemporal capital asset pricing and its implications for the Fama-French three factor model, unpublished paper, Anderson Graduate School of Management, UCLA.

Brown, Nerissa, Kesley Wei, and Russ Wermers, 2009, Analyst Recommendations, Mutual Fund Herding, and Overreaction in Stock Prices, University of Maryland working paper.

Brunnermeier, Marcus and Lasse Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201-2238.

Campbell, John Y., 1987, Stock returns and the term structure, *Journal of Financial Economics* 18, 373–399.

Campbell, John Y., 1991, A variance decomposition for stock returns, *Economic Journal* 101, 157–179.

Campbell, John Y., Christopher Polk, and Tuomo O. Vuolteenaho, 2010, Growth or glamour? Fundamentals and systematic risk in stock returns, *Review of Financial Studies* 23, 305-344.

Campbell, John Y. and Robert J. Shiller, 1988a, The dividend-price ratio and expectations of future dividends and discount factors, *Review of Financial Studies* 1, 195–228.

Campbell, John Y. and Robert J. Shiller, 1988b, Stock prices, earnings, and expected dividends, *Journal of Finance* 43, 661–676.

Campbell, John Y. and Robert J. Shiller, 2003, The long-run outlook for the US stock market: An update, forthcoming in Nicholas Barberis and Richard Thaler eds., *Advances in Behavioral Finance Vol. II*, Russell Sage Foundation, New York, NY.

Campbell, John Y. and Tuomo Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249–1275.

Chen, Joseph, Samuel Hanson, Harrison Hong, and Jeremy C. Stein, 2008, Do hedge funds profit from mutual-fund distress?, Harvard University working paper.

Carhart, M., 1997, On Persistence in Mutual Fund Performance, *Journal of Finance* 52, 56–82.

Chen, Huafeng, Shaojun Chen, and Feng Li, 2009, Firm-level comovement, UBC working paper.

Chevalier, Judith and Glen Ellison, 1997. Risk taking by mutual funds as a response to incentives, *Journal of Political Economy* 105, 1167–1200.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2003, The value spread, *Journal of Finance* 58, 609–641.

Cohen, Randolph B., Christopher Polk, and Tuomo Vuolteenaho, 2006, The price is (almost) right, unpublished paper, Harvard University and Northwestern University.

Coval, Josh and Eric Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 479–512.

Cremers, Martijn and Antti Petajisto, 2009, How active is your fund manager? A new measure that predicts performance, *Review of Financial Studies*.

DeMiguel, V., L. Garlappi and R. Uppal, 2007, Optimal versus Naive Diversi cation: How Inefficient Is the 1/N Portfolio Strategy?, *Review of Financial Studies,* forthcoming.

Eleswarapu, Venkat R. and Marc R. Reinganum, 2004, The predictability of aggregate stock market returns: Evidence based on glamour stocks, *Journal of Business* 77, 275–294.

Fama, Eugene F. & French, Kenneth R., 1988, Dividend yields and expected stock returns, *Journal of Financial Economics* 22, 3-25.

Fama, Eugene F. and Kenneth R. French, 1989, Business conditions and expected returns on stocks and bonds, *Journal of Financial Economics* 25, 23–49.

Fama, Eugene F. and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3–56.

Fama, E., MacBeth, J., 1973. Risk, return, and equilibrium: empirical tests, *Journal of Political Economy* 81, 607–636.

Greenwood, Robin and David Thesmar, 2009, Stock price fragility, HBS working paper.

Gromb, Denis and Dimitri Vayanos, 2002, Equilibrium and Welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361-407.

Ippolito, R., 1992. Consumer reaction to measures of poor quality: evidence from the mutual fund industry, *Journal of Law and Economics* 35, 45–70.

Keim, Donald and Robert Stambaugh, 1986, Predicting returns in the stock and bond markets, *Journal of Financial Economics* 17, 357-390.

Lou, Dong, 2009, A flow-based explanation for return predictability, London School of Economics working paper.

Merton, Robert C., On Estimating the Expected Return on the Market: An Exploratory Investigation, *Journal of Financial Economics* 8, 323-361.

Mitchell, Mark, Lasse Pedersen, and Todd Pulvino, 2007, Slow moving capital, *American Economic Review* 97, 215-220.

Pastor, Lubos and Robert Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.

Rozeff, Michael, 1984, Dividend yields are equity premiums, *Journal of Portfolio Management* 11, 68-75.

Sadka, Ronnie, 2006, Momentum and post-earnings-announcement drift anomalies: The role of liquidity risk, *Journal of Financial Economics* 80, 309-349.

Sadka, Ronnie, 2009, Liquidity risk and the cross-section of hedge-fund returns, forthcoming, *Journal of Financial Economics*.

Scholes, Myron, 1972, The market for corporate securities: Substitution versus price pressure and the effects of information on stock prices, *Journal of Business* 45, 179–211.

Shleifer, Andrei and Robert Vishny, 1992, Liquidation values and debt capacity: a market equilibrium approach, *Journal of Finance* 47, 1343–1366.

Shleifer, Andrei and Robert Vishny, 1997, The limits to arbitrage, *Journal of Finance* 52, 35–55.

Sirri, Eric, and Peter Tufano, 1998. Costly search and mutual fund flows, *Journal of Finance* 53, 1589–1622.

Sun, Zheng, 2008, Clustered institutional holdings and stock comovement, NYU working paper.

Vayanos, Dimitri, 2002, Flight to quality, flight to liquidity, and the pricing of risk, London School of Economics working paper.

Vayanos, Dimitri and Paul Woolley, 2008, An institutional theory of momentum and reversal, London School of Economics working paper.

Vuolteenaho, Tuomo, 2002, What drives firm-level stock returns, *Journal of Finance* 57, 233-264.

Table 1: **Number of Stocks, Analysts and Funds Per Year**

This table lists the total number of stocks, pairs of stocks, analysts and funds for every year of the sample period. The sample consists of all NYSE-AMEX-NASDAQ stocks that are above NYSE median capitalization as of the end of each quarter. Year t corresponds to the period June 1 of year $t-1$ to May 31 of year t. The number of unique stock pairs is $\frac{n*(n-1)}{2}$, where n is the number of stocks. The fourth column lists the number of analysts that cover (defined as issuing a one-year earnings forecast) at least one of the stocks in the sample. The fifth column lists the number of funds that hold at least one of the stocks in the sample.

Year	Stocks	Pairs	Analysts	Funds
1983	830	344035	1945	226
1984	824	339076	1987	236
1985	815	331705	1918	260
1986	798	318003	1873	314
1987	803	322003	1981	374
1988	767	293761	1820	400
1989	763	290703	1893	440
1990	801	320400	2110	477
1991	826	340725	1774	542
1992	845	356590	1649	618
1993	851	361675	1715	802
1994	864	372816	1868	922
1995	898	402753	2001	1015
1996	925	427350	2066	1124
1997	923	425503	2232	1280
1998	932	433846	2462	1457
1999	945	446040	2564	1592
2000	900	404550	2873	1742
2001	868	376278	2749	1875
2002	841	353220	2771	1919
2003	856	365940	2723	1914
2004	829	343206	2579	1909
2005	801	320400	2542	1874
2006	758	286903	2471	1754
2007	744	276396	2446	1693

Table 2: **Aggregate VAR**

The table shows the OLS parameter estimates for a first-order monthly aggregate VAR model including a constant, the log excess market return (r^e_M), the term yield spread (TY), the log price-earnings ratio (PE), and the small-stock value spread (VS). Each set of two rows corresponds to a different dependent variable. The first five columns report coefficients on the five explanatory variables and the sixth column reports the corresponding adjusted R^2. Standard errors are in parentheses. The sample period for the dependent variables is December 1928 - May 2009, providing 966 monthly data points.

Variable	Constant	$r^e_{M,t}$	TY_t	PE_t	VS_t	$\bar{R}^2$
$r^e_{M,t+1}$	0.0674	0.1118	0.0040	-0.0164	-0.0117	2.81%
	(0.0189)	(0.0318)	(0.0025)	(0.0048)	(0.0054)	
TY_{t+1}	-0.0278	0.0001	0.9212	-0.0051	0.0620	86.40%
	(0.0943)	(0.1585)	(0.0127)	(0.0243)	(0.0269)	
PE_{t+1}	0.0244	0.5181	0.0015	0.9923	-0.003	99.10%
	(0.0126)	(0.0212)	(0.0017)	(0.0032)	(0.0036)	
VS_{t+1}	0.0180	0.0045	0.0008	-0.0010	0.9903	98.24%
	(0.0169)	(0.0283)	(0.0022)	(0.0043)	(0.0048)	

Table 3: **Firm-level VAR**

The table shows the pooled-WLS parameter estimates for a first-order monthly firm-level VAR model. The model state vector includes the log stock return (r), stock momentum (MOM), and the log book-to-market (BM). We define MOM as the cumulative stock return over the last year, but excluding the most recent month. All three variables are market-adjusted: r is adjusted by subtracting r_M while MOM and BM are adjusted by removing the respective month-specific cross-sectional means. Rows corresponds to dependent variables and columns to independent (lagged dependent) variables. The first three columns report coefficients on the three explanatory variables and the fourth column reports the corresponding adjusted R^2. The weights used in the WLS estimation are proportional to the inverse of the number of stocks in the corresponding cross section. Standard errors (in parentheses) take into account clustering in each cross section. The sample period for the dependent variables is January 1954 - December 2008, providing 660 monthly cross-sections and 1,658,049 firm-months.

Variable	$r_{i,t}$	$MOM_{i,t}$	$BM_{i,t}$	R^2
$r_{i,t+1}$	-0.0470	0.0206	0.0048	0.64%
	(0.0066)	(0.0023)	(0.0007)	
$MOM_{i,t+1}$	0.9555	0.9051	-0.0015	91.85%
	(0.0052)	(0.0018)	(0.0007)	
$BM_{i,t+1}$	0.0475	-0.0107	0.9863	97.10%
	(0.0050)	(0.0017)	(0.0011)	

Table 4: **Ownership, Coverage, and Stock Returns: Summary Statistics**

This table reports summary statistics for the sample defined in Table 1 over the following variables: number of analysts that cover each stock, number of stocks covered by each analyst, number of funds that hold each stock and number of stocks held by each fund. We also report summary statistics for the net monthly stock return ($R_{i,t}$), cash flow news ($N_{CF,i,t}$), discount rate news ($N_{DR,i,t}$) as well as the cross products of net monthly returns and their components. There are a total of 420,108 analyst-months and 297,312 fund-months. There are 16,785 stock-years and 7,075,067 pair-years. Summary statistics are reported for those observations for which values of all variables are available. Panel A reports these summary statistics for the full sample, while Panels B, C, and D report summary statistics for the sample by decade.

PANEL A: 1983-2007

Variable	Mean	Median	Std	Min	Max
Analysts per Stock	17.8	16	10.2	1	68
Stocks per Analyst	6.9	5	7.3	1	95
Funds per Stock	63.8	37	78.9	1	799
Stocks per Fund	55.1	40	61.8	1	1026
$R_{i,t}$	0.0113	0.0102	0.1040	-0.9968	2.2663
$-N_{DR,i,t}$	0.0039	0.0049	0.0539	-0.9106	0.7997
$N_{CF,i,t}$	-0.0033	-0.0021	0.0855	-2.2437	1.2282
$R_{i,t}R_{j,t}$	0.0023	0.0002	0.0102	-1.1332	4.6802
$R_{i,t}R_{i,t}$	0.0109	0.0028	0.0365	0.0000	5.1363
$N_{DR,i,t}N_{DR,j,t}$	0.0022	0.0006	0.0015	-0.6131	0.4112
$N_{CF,i,t}N_{CF,j,t}$	0.0007	0.0001	0.0071	-1.1618	2.2651
$-N_{CF,i,t}N_{DR,j,t}$	-0.0011	-0.0003	0.0056	-1.7364	1.6953

PANEL B: 1983-1989

Variable	Mean	Median	Std	Min	Max
Analysts per Stock	19.6	18	12.2	1	63
Stocks per Analyst	8.6	6	9.4	1	95
Funds per Stock	13.4	9	13.7	1	164
Stocks per Fund	39.9	32	32.9	1	433
$R_{i,t}$	0.0159	0.0128	0.0931	-0.7614	1.3564
$-N_{DR,i,t}$	0.0010	0.0003	0.0529	-0.6545	0.7997
$N_{CF,i,t}$	-0.0050	-0.0053	0.0699	-1.0319	0.8077
$R_{i,t}R_{j,t}$	0.0026	0.0002	0.0081	-0.3457	1.1692
$R_{i,t}R_{i,t}$	0.0089	0.0027	0.0228	0.0000	1.8398
$N_{DR,i,t}N_{DR,j,t}$	0.0022	0.0007	0.0013	-0.2385	0.1915
$N_{CF,i,t}N_{CF,j,t}$	0.0005	0.0000	0.0048	-0.3045	0.6535
$-N_{CF,i,t}N_{DR,j,t}$	-0.0008	-0.0003	0.0045	-0.4420	0.5010

PANEL C: 1990-1999

Variable	Mean	Median	Std	Min	Max
Analysts per Stock	17.3	16	9.4	1	68
Stocks per Analyst	7.4	5	7.8	1	95
Funds per Stock	55.1	40	54.2	1	583
Stocks per Fund	51.8	39	56.9	1	820
$R_{i,t}$	0.0138	0.0111	0.1045	-0.8265	2.2663
$-N_{DR,i,t}$	0.0131	0.0121	0.0478	-0.5696	0.6107
$N_{CF,i,t}$	-0.0060	-0.0044	0.0862	-1.2374	1.2282
$R_{i,t}R_{j,t}$	0.0019	0.0002	0.0105	-1.1332	4.6802
$R_{i,t}R_{i,t}$	0.0111	0.0029	0.0415	0.0000	5.1363
$N_{DR,i,t}N_{DR,j,t}$	0.0018	0.0004	0.0014	-0.2125	0.3580
$N_{CF,i,t}N_{CF,j,t}$	0.0006	0.0000	0.0072	-0.6511	1.3763
$-N_{CF,i,t}N_{DR,j,t}$	-0.0009	-0.0002	0.0052	-0.7384	0.5000

PANEL D: 2000-2007

Variable	Mean	Median	Std	Min	Max
Analysts per Stock	16.9	16	9.0	1	62
Stocks per Analyst	5.4	4	4.6	1	65
Funds per Stock	129.1	102	98.0	1	799
Stocks per Fund	59.7	43	67.6	1	1026
$R_{i,t}$	0.0032	0.0065	0.1140	-0.9968	1.5625
$-N_{DR,i,t}$	-0.0039	0.0004	0.0602	-0.9106	0.6733
$N_{CF,i,t}$	0.0019	0.0052	0.0994	-2.2437	1.1418
$R_{i,t}R_{j,t}$	0.0023	0.0001	0.0122	-1.0351	2.2124
$R_{i,t}R_{i,t}$	0.0130	0.0029	0.0421	0.0000	2.4414
$N_{DR,i,t}N_{DR,j,t}$	0.0027	0.0006	0.0019	-0.6131	0.4112
$N_{CF,i,t}N_{CF,j,t}$	0.0010	0.0001	0.0094	-1.1618	2.2651
$-N_{CF,i,t}N_{DR,j,t}$	-0.0017	-0.0007	0.0073	-1.7364	1.6953

Table 5: **The Cross-sectional Distribution of Common Fund Ownership**

This table reports the distribution of the variable $F_{ij,t}$ measuring the number of funds holding both stocks in a pair over the last quarter. There are 7,075,067 pair-years.

FUNDS IN COMMON ($F_{ij,t}$)			Percentiles						
Year	Mean	Std	0%	25%	50%	75%	95%	99%	100%
ALL	9.26	16.97	0	1	3	11	37	76	640
1983	0.74	1.46	0	0	0	1	3	7	52
1984	0.77	1.55	0	0	0	1	3	7	56
1985	0.89	1.77	0	0	0	1	4	8	58
1986	1.07	2.10	0	0	0	1	5	10	62
1987	1.68	3.05	0	0	1	2	7	14	97
1988	2.38	3.64	0	0	1	3	9	17	80
1989	2.56	3.98	0	0	1	3	9	19	82
1990	2.87	4.63	0	0	1	4	11	21	115
1991	3.53	5.20	0	0	2	5	13	24	129
1992	4.30	6.12	0	0	2	7	15	27	132
1993	6.52	7.75	0	2	4	9	21	36	153
1994	7.39	8.70	0	2	4	10	23	41	186
1995	8.14	10.38	0	2	5	11	26	49	231
1996	7.97	10.56	0	2	4	10	26	50	280
1997	9.34	12.25	0	3	5	12	30	58	288
1998	10.38	15.03	0	3	6	13	33	72	394
1999	12.46	18.13	0	3	7	16	39	84	499
2000	14.86	21.89	0	4	8	19	47	106	543
2001	19.76	24.75	0	6	12	25	59	121	577
2002	22.05	26.92	0	7	13	28	64	131	615
2003	22.55	26.64	0	7	14	29	65	129	630
2004	22.36	24.92	0	8	15	29	64	121	571
2005	22.80	24.35	0	8	15	29	64	120	500
2006	22.19	23.06	0	8	15	28	62	115	443
2007	25.73	23.51	0	12	19	32	66	121	463

Table 6: **The Cross-sectional Distribution of Common Analyst Coverage**

This table reports the distribution of the variable $A_{ij,t}$ measuring the number of analysts forecasting one-year EPS for both stocks in a pair over the past quarter. There are 7,075,067 pair-years.

ANALYSTS IN COMMON ($A_{ij,t}$)			Percentiles						
Year	Mean	Std	0%	25%	50%	75%	95%	99%	100%
ALL	0.24	1.46	0	0	0	0	1	6	53
1983	0.38	1.73	0	0	0	0	2	8	43
1984	0.43	1.96	0	0	0	0	2	10	47
1985	0.42	1.86	0	0	0	0	2	9	48
1986	0.44	1.98	0	0	0	0	2	10	49
1987	0.48	2.32	0	0	0	0	2	13	46
1988	0.48	2.24	0	0	0	0	2	13	47
1989	0.45	2.19	0	0	0	0	2	13	45
1990	0.39	1.97	0	0	0	0	1	10	53
1991	0.35	1.71	0	0	0	0	1	9	39
1992	0.31	1.57	0	0	0	0	1	8	35
1993	0.32	1.75	0	0	0	0	1	9	42
1994	0.31	1.66	0	0	0	0	1	8	39
1995	0.25	1.41	0	0	0	0	1	7	39
1996	0.23	1.34	0	0	0	0	1	6	38
1997	0.21	1.20	0	0	0	0	1	5	44
1998	0.18	1.15	0	0	0	0	1	4	41
1999	0.17	1.10	0	0	0	0	1	4	39
2000	0.16	1.07	0	0	0	0	1	4	40
2001	0.15	1.01	0	0	0	0	1	4	35
2002	0.14	1.05	0	0	0	0	0	4	41
2003	0.15	1.13	0	0	0	0	1	4	44
2004	0.16	1.17	0	0	0	0	1	4	44
2005	0.16	1.20	0	0	0	0	0	5	43
2006	0.17	1.24	0	0	0	0	0	5	44
2007	0.16	1.18	0	0	0	0	0	5	37

Table 7: **Connected Comovement**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, for the sample of stocks defined in Table 1. We estimate

$$\begin{aligned} r_{i,t+1}r_{j,t+1} &= a + b_f{*}F^*_{ij,t}+b_a{*}A^*_{ij,t}+b_s{*}DIFF\_SIZE^*_{ij,t}+b_b{*}DIFF\_BEME^*_{ij,t} \\ &+b_m * DIFF\_MOM^*_{ij,t} + b_k{*}NUM\_SIC^*_{ij,t}+b_{s1}{*}SIZE1^*_{ij,t} \\ &+b_{s2}{*}SIZE2^*_{ij,t}+b_{s12}{*}SIZE1SIZE2^*_{ij,t}+\varepsilon_{ij,t} \end{aligned}$$

The independent variables are updated annually and include our main measures of institutional connectedness, common funds ($F_{ij,t}$) and common analysts ($A_{ij,t}$), and a series of controls at time t. We assign stocks to size, BE/ME and momentum deciles and measure the difference in ranking across the two stocks in the pair ($DIFF\_SIZE_{ij,t}$, $DIFF\_BEME_{ij,t}$, and $DIFF\_MOM_{ij,t}$ respectively). We also measure the number of similar SIC digits, $NUM\_SIC_{ij,t}$, for the two stocks in a pair as well as the size percentile of each stock in the pair and an interaction ($SIZE1_{ij,t}$, $SIZE2_{ij,t}$, and $SIZE1SIZE2_{ij,t}$ where stock 1 is always the larger stock in the pair). All independent variables are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. We report estimates of regressions using various subsets of these variables in Panel A. For regression (5), we replace the variables measuring the difference in size, BE/ME, and momentum deciles as well as the similarity in SIC code across the pair with a full set of dummy variables, which we report in Panel B. We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

PANEL A

Dependent Variable: $r_{i,t+1}r_{j,t+1}$

	(1)	(2)	(3)	(4)	(5)
$F^*_{ij,t}$	0.00030	0.00027	0.00024	0.00050	0.00050
	(6.11)	(5.73)	(5.64)	(6.77)	(6.80)
$A^*_{ij,t}$		0.00018	0.00010	0.00013	0.00011
		(7.49)	(6.20)	(7.87)	(9.59)
Constant	0.00216	0.00216	0.00216	0.00217	0.00355
	(8.46)	(8.46)	(8.46)	(8.47)	(7.89)
$DIFF\_SIZE^*_{ij,t}$			0.00002	-0.00028	
			(1.17)	(-4.77)	
$DIFF\_BEME^*_{ij,t}$			0.00012	0.00009	
			(2.78)	(2.30)	
$DIFF\_MOM^*_{ij,t}$			0.00012	0.00012	
			(2.28)	(2.37)	
$NUM\_SIC^*_{ij,t}$			0.00020	0.00019	
			(7.30)	(7.02)	
$SIZE1^*_{ij,t}$				0.00097	0.00075
				(5.51)	(5.76)
$SIZE2^*_{ij,t}$				0.00013	0.00030
				(2.30)	(4.25)
$SIZE1SIZE2^*_{ij,t}$				-0.00057	-0.00054
				(-4.79)	(-4.72)

PANEL B

dummy estimates for specification (5) in Panel A				
Variable Value	$DIFF\_SIZE_{ij,t}$	$DIFF\_BEME_{ij,t}$	$DIFF\_MOM_{ij,t}$	$NUM\_SIC_{ij,t}$
0				-0.00105
				(-3.56)
1	0.00003	-0.00010	-0.00028	-0.00062
	(2.34)	(-4.03)	(-6.02)	(-2.24)
2	0.00011	-0.00012	-0.00042	-0.00078
	(3.21)	(-3.26)	(-5.47)	(-3.55)
3	0.00019	-0.00017	-0.00048	0.00040
	(3.48)	(-3.14)	(-5.38)	(2.20)
4	0.00025	-0.00022	-0.00052	
	(3.50)	(-3.28)	(-5.09)	
5	0.00028	-0.00025	-0.00055	
	(3.18)	(-3.12)	(-4.67)	
6	0.00028	-0.00028	-0.00055	
	(2.76)	(-2.95)	(-4.21)	
7	0.00028	-0.00033	-0.00052	
	(2.32)	(-2.90)	(-3.43)	
8	0.00025	-0.00039	-0.00044	
	(1.82)	(-2.69)	(-2.29)	
9	0.00021	-0.00039	-0.00013	
	(1.29)	(-2.12)	(-0.52)	

Table 8: **Connected Comovement: Alternative Measures**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting measures of stock-pair comovement for the sample of stocks defined in Table 1. In particular, we forecast the realized cross-product of monthly returns, $r_{i,t+1}r_{j,t+1}$, the corrected sum of squares ($S_{r_ir_j}$) using daily return data in month t+1, as well as the daily return Fisher correlation (ρ_{Fisher}) or the daily return Pearson correlation ($\rho_{Pearson}$) realized in month t+1. We estimate

$$\begin{aligned}\mathbf{y} &= a + b_f{*}F^*_{ij,t}{+}b_a{*}A^*_{ij,t}{+}b_s{*}DIFF\_SIZE^*_{ij,t}{+}b_b{*}DIFF\_BEME^*_{ij,t}\\ &\quad +b_m * DIFF\_MOM^*_{ij,t} + b_k{*}NUM\_SIC^*_{ij,t}{+}b_{s1}{*}SIZE1^*_{ij,t}\\ &\quad +b_{s2}{*}SIZE2^*_{ij,t}{+}b_{s12}{*}SIZE1SIZE2^*_{ij,t}{+}\varepsilon_{ij,t}\end{aligned}$$

where $\mathbf{y}= [r_{i,t+1}r_{j,t+1}, S_{r_ir_j}, \rho_{Fisher}, \rho_{Pearson}]$. The independent variables are updated quarterly and include our main measures of institutional connectedness, common funds ($F_{ij,t}$) and common analysts ($A_{ij,t}$), and a series of controls at time t. We assign stocks to size, BE/ME and momentum deciles and measure the difference in ranking across the two stocks in the pair ($DIFF\_SIZE_{ij,t}$, $DIFF\_BEME_{ij,t}$, and $DIFF\_MOM_{ij,t}$ respectively). We also measure the number of similar SIC digits, $NUM\_SIC_{ij,t}$, for the two stocks in a pair as well as the size percentile of each stock in the pair and an interaction ($SIZE1_{ij,t}$, $SIZE2_{ij,t}$, and $SIZE1SIZE2_{ij,t}$). All of these variables are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

Variable	$r_{i,t+1}r_{j,t+1}$	S_{xy}	$\rho_{Pearson}$	ρ_{Fisher}
$F^*_{ij,t}$	0.00050	0.00037	0.01806	0.02020
	(6.77)	(9.05)	(16.34)	(16.10)
$A^*_{ij,t}$	0.00013	0.00010	0.01269	0.01605
	(7.87)	(5.89)	(13.64)	(12.77)
Constant	0.00217	0.00185	0.18278	0.20026
	(8.47)	(8.17)	(20.93)	(19.74)
$DIFF\_SIZE^*_{ij,t}$	-0.00028	-0.00007	0.00925	0.01143
	(-4.77)	(-1.64)	(6.72)	(7.36)
$DIFF\_BEME^*_{ij,t}$	0.00009	0.00001	0.00264	0.00319
	(2.30)	(0.85)	(5.53)	(5.75)
$DIFF\_MOM^*_{ij,t}$	0.00012	-0.00000	0.00615	0.00724
	(2.37)	(-0.30)	(8.66)	(8.58)
$NUM\_SIC^*_{ij,t}$	0.00019	0.00014	0.00909	0.01096
	(7.02)	(4.88)	(11.99)	(11.59)
$SIZE1^*_{ij,t}$	0.00097	0.00025	-0.03347	-0.04032
	(5.51)	(2.60)	(-8.07)	(-8.44)
$SIZE2^*_{ij,t}$	0.00013	0.00007	-0.00582	-0.00634
	(2.30)	(1.34)	(-2.99)	(-2.88)
$SIZE1SIZE2^*_{ij,t}$	-0.00057	-0.00019	0.02160	0.02636
	(-4.79)	(-2.82)	(7.80)	(8.17)

Table 9: **Cross-sectional Variation in Connected Comovement: Stock Size and Fund Flows**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, for the sample of stocks defined in Table 1. We estimate

$$\begin{aligned} r_{i,t+1}r_{j,t+1} &= a+\sum_{k=1}^{5}\sum_{l=1}^{5} b_{f-k,l}*F^*_{ij,t}+b_a*A^*_{ij,t}+b_s*DIFF\_SIZE^*_{ij,t}+b_b*DIFF\_BEME^*_{ij,t} \\ &+b_m*DIFF\_MOM^*_{ij,t}+b_k*NUM\_SIC^*_{ij,t}+b_{s1}*SIZE1^*_{ij,t} \\ &+b_{s2}*SIZE2^*_{ij,t}+b_{s12}*SIZE1SIZE2^*_{ij,t}+\varepsilon_{ij,t} \end{aligned}$$

The independent variables are updated annually and include our main measures of institutional connectedness, common funds ($F_{ij,t}$) and common analysts ($A_{ij,t}$), and a series of controls at time t. We assign stocks to size, BE/ME and momentum deciles and measure the difference in ranking across the two stocks in the pair ($DIFF\_SIZE_{ij,t}$, $DIFF\_BEME_{ij,t}$, and $DIFF\_MOM_{ij,t}$ respectively). We also measure the number of similar SIC digits, $NUM\_SIC_{ij,t}$, for the two stocks in a pair as well as the size percentile of each stock in the pair and an interaction ($SIZE1_{ij,t}$, $SIZE2_{ij,t}$, and $SIZE1SIZE2_{ij,t}$). All of these variables are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. The specification is similar to regression (4) in Table 7, but we interact the common fund variable with dummies for the ranking of the pair based on quarterly independent sorts on the pair's total market capitalization (k dimension of $b_{f-k,l}$) and the total fund flows of the common funds (l dimension of $b_{f-k,l}$). We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

Dependent variable: $r_{i,t+1}r_{j,t+1}$ (Controls included but not reported)						
$b_{f-k,l}$ estimates		Size of the pair (k)				
		Low	2	3	4	High
Total net flow from common funds	Low	0.00093	0.00084	0.00068	0.00053	0.00047
		(4.30)	(5.66)	(6.32)	(6.11)	(5.29)
	2	0.00072	0.00067	0.00058	0.00046	0.00043
		(5.97)	(6.21)	(6.00)	(5.51)	(5.13)
	3	0.00074	0.00068	0.00062	0.00051	0.00046
		(5.08)	(5.25)	(5.25)	(5.09)	(4.62)
	4	0.00069	0.00061	0.00057	0.00042	0.00036
		(6.50)	(6.76)	(7.39)	(6.61)	(5.43)
	High	0.00124	0.00103	0.00081	0.00064	0.00050
		(6.54)	(6.49)	(7.36)	(6.72)	(6.02)

Table 10: **Understanding Connected Comovement: A Decomposition**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, as well as the cross products of the return components (cash-flow-news and discount-rate-news), $N_{CF,i,t+1}N_{CF,j,t+1}$, $-N_{DR,j,t+1}N_{CF,i,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$, and $N_{DR,i,t+1}N_{DR,j,t+1}$ for the sample of stocks defined in Table 1. We estimate

$$\begin{aligned} y &= a + b_f{*}F^*_{ij,t}{+}b_a{*}A^*_{ij,t}{+}b_s{*}DIFF\_SIZE^*_{ij,t}{+}b_b{*}DIFF\_BEME^*_{ij,t} \\ &\quad +b_m * DIFF\_MOM^*_{ij,t} + b_k{*}NUM\_SIC^*_{ij,t}{+}b_{s1}{*}SIZE1^*_{ij,t} \\ &\quad +b_{s2}{*}SIZE2^*_{ij,t}{+}b_{s12}{*}SIZE1SIZE2^*_{ij,t}{+}\varepsilon_{ij,t} \end{aligned}$$

where $y = [r_{i,t+1}r_{j,t+1};\ N_{CF,i,t+1}N_{CF,j,t+1};\ -N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1};\ N_{DR,i,t+1}N_{DR,j,t+1}]$. The independent variables are updated quarterly and include our main measures of institutional connectedness, common funds ($F_{ij,t}$) and common analysts ($A_{ij,t}$), and a series of controls at time t. We assign stocks to size, BE/ME and momentum deciles and measure the difference in ranking across the two stocks in the pair ($DIFF\_SIZE_{ij,t}$, $DIFF\_BEME_{ij,t}$, and $DIFF\_MOM_{ij,t}$ respectively). We also measure the number of similar SIC digits, $NUM\_SIC_{ij,t}$, for the two stocks in a pair as well as the size percentile of each stock in the pair and an interaction ($SIZE1_{ij,t}$, $SIZE2_{ij,t}$, and $SIZE1SIZE2_{ij,t}$). All of these variables are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. The return components are constructed from the aggregate and firm-level VARs estimated in Tables 2 and 3 as described in the Appendix. We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

Variable	$r_{i,t+1}r_{j,t+1}$	$N_{CF,i}N_{CF,j}$	$-N_{DR,i}N_{CF,j}+$ $-N_{DR,j}N_{CF,i}$	$N_{DR,i}N_{DR,j}$
	(1)	(2)	(3)	(4)
$F^*_{ij,t}$	0.00050	0.00018	0.00017	0.00003
	(6.77)	(3.66)	(4.27)	(3.51)
$A^*_{ij,t}$	0.00013	0.00015	-0.00005	0.00001
	(7.87)	(7.34)	(-3.65)	(3.43)
Constant	0.00217	0.00069	-0.00112	0.00209
	(8.47)	(7.48)	(-5.37)	(9.44)
$DIFF\_SIZE^*_{ij,t}$	-0.00028	-0.00016	-0.00009	-0.00001
	(-4.77)	(-4.04)	(-1.53)	(-0.80)
$DIFF\_BEME^*_{ij,t}$	0.00009	0.00012	-0.00007	0.00003
	(2.30)	(3.81)	(-3.47)	(5.59)
$DIFF\_MOM^*_{ij,t}$	0.00012	0.00021	-0.00013	0.00001
	(2.37)	(4.16)	(-5.51)	(1.77)
$NUM\_SIC^*_{ij,t}$	0.00019	0.00011	0.00003	0.00001
	(7.02)	(8.45)	(2.43)	(4.51)
$SIZE1^*_{ij,t}$	0.00097	0.00053	0.00029	0.00004
	(5.51)	(4.34)	(1.83)	(1.21)
$SIZE2^*_{ij,t}$	0.00013	0.00003	0.00001	0.00001
	(2.30)	(0.85)	(0.26)	(1.44)
$SIZE1SIZE2^*_{ij,t}$	-0.00057	-0.00031	-0.00020	-0.00003
	(-4.79)	(-3.47)	(-1.83)	(-1.38)

Table 11: **Connected Comovement: Cross-sectional Variation in the Decomposition**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, as well as the realized cross products of the return components (cash-flow-news and discount-rate-news), $N_{CF,i,t+1}N_{CF,j,t+1}$, $-N_{DR,j,t+1}N_{CF,i,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$, and $N_{DR,i,t+1}N_{DR,j,t+1}$ for the sample of stocks defined in Table 1. We estimate

$$\begin{aligned} y &= a+\sum_{k=1}^{5}\sum_{l=1}^{5} b_{f-k,l} * F^*_{ij,t} + b_a * A^*_{ij,t} + b_s * DIFF\_SIZE^*_{ij,t} + b_b * DIFF\_BEME^*_{ij,t} \\ &\quad + b_m * DIFF\_MOM^*_{ij,t} + b_k * NUM\_SIC^*_{ij,t} + b_{s1} * SIZE1^*_{ij,t} \\ &\quad + b_{s2} * SIZE2^*_{ij,t} + b_{s12} * SIZE1SIZE2^*_{ij,t} + \varepsilon_{ij,t} \end{aligned}$$

where $y = N_{CF,i,t+1}N_{CF,j,t+1}$ [Panel A], $-N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$ [Panel B], and $N_{DR,i,t+1}N_{DR,j,t+1}$ [Panel C]. The independent variables are updated annually and include our main measures of institutional connectedness, common funds ($F_{ij,t}$) and common analysts ($A_{ij,t}$), and a series of controls at time t. We assign stocks to size, BE/ME and momentum deciles and measure the difference in ranking across the two stocks in the pair ($DIFF\_SIZE_{ij,t}$, $DIFF\_BEME_{ij,t}$, and $DIFF\_MOM_{ij,t}$ respectively). We also measure the number of similar SIC digits, $NUM\_SIC_{ij,t}$, for the two stocks in a pair as well as the size percentile of each stock in the pair and an interaction ($SIZE1_{ij,t}$, $SIZE2_{ij,t}$, and $SIZE1SIZE2_{ij,t}$). All of these variables are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. The return components are constructed from the aggregate and firm-level VARs estimated in Tables 2 and 3. The specifications are similar to those in Table 10, but we interact the common fund variable with dummies for the ranking of the pair based on annual independent sorts on the pair's total market capitalization (k dimension of $b_{f-k,l}$) and the total fund flows of the common funds (l dimension of $b_{f-k,l}$). We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

PANEL A: Dependent var: $N_{CF,i,t+1}N_{CF,j,t+1}$						
Controls included but not shown						
$b_{f-k,l}$ estimates		Size of the pair (k)				
		Low	2	3	4	High
	Low	0.00034	0.00031	0.00028	0.00028	0.00022
		(4.79)	(4.98)	(4.16)	(3.42)	(2.72)
Total	2	0.00018	0.00018	0.00016	0.00015	0.00012
net		(4.03)	(4.85)	(4.18)	(3.70)	(2.43)
flow	3	0.00021	0.00018	0.00012	0.00015	0.00010
from		(3.70)	(3.23)	(2.18)	(2.57)	(1.71)
common	4	0.00024	0.00023	0.00020	0.00020	0.00013
funds (l)		(4.71)	(4.18)	(4.03)	(3.78)	(2.60)
	High	0.00042	0.00028	0.00025	0.00021	0.00011
		(5.16)	(5.13)	(5.30)	(4.70)	(2.79)

PANEL B: Dependent var: $-N_{DR,i,t+1}N_{CF,j,t+1} - N_{DR,j,t+1}N_{CF,i,t+1}$						
Controls included but not shown						
$b_{f-k,l}$ estimates		Size of the pair (k)				
		Low	2	3	4	High
	Low	0.00030	0.00029	0.00020	0.00011	0.00014
		(3.13)	(3.65)	(2.53)	(1.30)	(1.66)
Total	2	0.00033	0.00029	0.00025	0.00016	0.00019
net		(4.49)	(4.13)	(3.90)	(2.82)	(3.19)
flow	3	0.00032	0.00031	0.00033	0.00021	0.00023
from		(3.31)	(3.70)	(3.86)	(3.20)	(3.53)
common	4	0.00027	0.00022	0.00022	0.00011	0.00015
funds (l)		(3.86)	(3.72)	(3.99)	(2.09)	(2.60)
	High	0.00049	0.00048	0.00034	0.00025	0.00026
		(4.57)	(4.73)	(5.03)	(4.01)	(4.69)

PANEL C: Dependent var: $N_{DR,i,t+1}N_{DR,j,t+1}$						
Controls included but not shown						
$b_{f-k,l}$ estimates		Size of the pair (k)				
		Low	2	3	4	High
	Low	0.00005	0.00004	0.00005	0.00004	0.00004
		(2.54)	(2.44)	(3.07)	(2.57)	(2.63)
Total	2	0.00004	0.00004	0.00004	0.00003	0.00003
net		(2.52)	(3.09)	(3.00)	(2.68)	(2.32)
flow	3	0.00006	0.00004	0.00004	0.00004	0.00004
from		(2.68)	(2.65)	(2.60)	(3.17)	(2.64)
common	4	0.00003	0.00002	0.00003	0.00002	0.00002
funds (l)		(2.20)	(2.11)	(2.48)	(2.26)	(2.32)
	High	0.00005	0.00005	0.00004	0.00004	0.00004
		(2.38)	(2.61)	(3.09)	(2.77)	(2.77)

Table 12: **Connected Comovement: Additional Controls**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, the daily return Fisher correlation (ρ_{Fisher}), and the cross products of the return components (cash-flow-news and discount-rate-news), $N_{CF,i,t+1}N_{CF,j,t+1}$, $-N_{DR,j,t+1}N_{CF,i,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$, and $N_{DR,i,t+1}N_{DR,j,t+1}$ for the sample of stocks defined in Table 1. We estimate

$$\begin{aligned}\mathbf{y} \;=\; & a + b_f{*}F^*_{ij,t}{+}b_a{*}A^*_{ij,t}{+}b_s{*}DIFF\_SIZE^*_{ij,t}{+}b_b{*}DIFF\_BEME^*_{ij,t} \\ & +b_m * DIFF\_MOM^*_{ij,t} + b_k{*}NUM\_SIC^*_{ij,t}{+}b_{s1}{*}SIZE1^*_{ij,t} \\ & +b_{s2}{*}SIZE2^*_{ij,t}{+}b_{s12}{*}SIZE1SIZE2^*_{ij,t}{+}b_{ret} * RETCORR_{ij,t} \\ & +b_{roe} * ROECORR_{ij,t}{+} + b_{vol} * VOLCORR_{ij,t} + b_{grth} * DIFFGRTH_{ij,t} \\ & +b_{lev}{*}DIFFLEV_{ij,t} + b_{state} * D_{STATE_{ij,t}}{+}b_{index} * D_{INDEXij,t} \\ & +b_{price} * DIFFPRICE_{ij,t}{+}b_{listing} * D_{LISTING_{ij,t}} + \varepsilon_{ij,t}\end{aligned}$$

where $y = [r_{i,t+1}r_{j,t+1};\, \rho_{Fisher};N_{CF,i,t+1}N_{CF,j,t+1};\, -N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1};\, N_{DR,i,t+1}N_{DR,j,t+1}]$. This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, the daily return Fisher correlation (ρ_{Fisher}) as well as the cross products of the return components (cash-flow-news and discount-rate-news), $(N_{CF,i,t+1}) * (N_{CF,j,t+1})$, $(N_{CF,i,t+1}) * (-N_{DR,j,t+1})$ and $(-N_{DR,i,t+1}) * (-N_{DR,j,t+1})$ for the sample of stocks defined in Table 1. We estimate the same equation as in Table 8, but with additional variables as a robustness check. The additional variables are constructed as in Chen, Chen, Li (2009) and are as follows: past return correlation, $RETCORR_{ij,t}$; past profitability correlation, $ROECORR_{ij,t}$; the past correlation in the stocks abnormal trading volume, $VOLCORR_{ij,t}$, the absolute value of the difference in five-year log sales growth rates, $DIFFGRTH_{ij,t}$; the absolute difference in financial leverage ratios (defined as long-term debt / total assets), $DIFFLEV_{ij,t}$; the absolute value of the difference in the two stocks' log share prices, $DIFFPRICE_{ij,t}$; a dummy variable in the two firms are located in the same state; $D_{STATE_{ij,t}}$; a dummy variable if the two stocks belong to the same S&P major, mid-cap, or small-cap index, $D_{INDEXij,t}$; and a dummy variable if the two stocks are on the same stock exchange, $D_{LISTING_{ij,t}}$. All of these variables (except the dummies) are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. The return components are constructed from the aggregate and firm-level VARs estimated in Tables 2 and 3 as described in the Appendix. We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

Variable	$r_{i,t+1}r_{j,t+1}$	ρ_{Fisher}	$N_{CF,i}N_{CF,j}$	$-N_{DR,i}N_{CF,j}$ $-N_{DR,j}N_{CF,i}$	$N_{DR,i}N_{DR,j}$
$F^*_{ij,t}$	0.00051	0.01080	0.00010	0.00027	0.00002
	(6.44)	(11.80)	(2.82)	(5.59)	(2.05)
$A^*_{ij,t}$	0.00008	0.01336	0.00011	-0.00004	0.00000
	(5.18)	(11.01)	(8.67)	(-3.87)	(0.94)
Constant	0.00228	0.19159	0.00051	-0.00076	0.00203
	(8.28)	(17.09)	(6.42)	(-4.42)	(8.94)
$DIFF\_SIZE^*_{ij,t}$	-0.00023	0.01430	-0.00013	-0.00007	-0.00001
	(-4.00)	(9.10)	(-3.75)	(-1.30)	(-0.88)
$DIFF\_BEME^*_{ij,t}$	0.00006	0.00189	0.00007	-0.00004	0.00002
	(1.94)	(4.13)	(3.75)	(-2.94)	(5.27)
$DIFF\_MOM^*_{ij,t}$	0.00007	0.00456	0.00015	-0.00009	0.00000
	(1.74)	(6.70)	(3.95)	(-5.80)	(0.05)
$NUM\_SIC^*_{ij,t}$	0.00013	0.00846	0.00008	0.00002	0.00000
	(5.47)	(9.74)	(8.38)	(1.28)	(2.16)
$SIZE1^*_{ij,t}$	0.00081	-0.04500	0.00044	0.00024	0.00005
	(4.81)	(-9.11)	(4.28)	(1.58)	(1.25)
$SIZE2^*_{ij,t}$	0.00012	-0.00184	0.00002	0.00002	0.00001
	(2.37)	(-0.90)	(0.63)	(0.51)	(1.06)
$SIZE1SIZE2^*_{ij,t}$	-0.00048	0.02815	-0.00024	-0.00018	-0.00003
	(-4.28)	(8.46)	(-3.46)	(-1.76)	(-1.37)
$RETCORR^*_{ij,t}$	0.00040	0.02369	0.00026	0.00002	0.00004
	(8.02)	(13.57)	(4.44)	(0.51)	(4.82)
$ROECORR^*_{ij,t}$	0.00005	0.00116	0.00002	0.00002	0.00000
	(3.67)	(3.42)	(3.24)	(2.71)	(1.10)
$VOLCORR^*_{ij,t}$	0.00005	0.00389	0.00003	0.00001	0.00000
	(3.99)	(7.12)	(3.32)	(0.95)	(0.35)
$DIFFGRTH^*_{ij,t}$	0.00016	-0.00217	-0.00006	0.00020	-0.00001
	(5.50)	(-2.76)	(-3.01)	(5.95)	(-2.13)
$DIFFLEV^*_{ij,t}$	-0.00002	-0.00319	-0.00000	-0.00001	0.00000
	(-1.40)	(-6.39)	(-0.18)	(-1.25)	(1.79)
$DIFFPRICE^*_{ij,t}$	0.00007	-0.00592	-0.00002	0.00007	0.00000
	(3.61)	(-9.55)	(-1.89)	(3.88)	(0.84)
$D_{STATE_{ij,t}}$	0.00049	0.00864	0.00010	0.00029	0.00000
	(5.80)	(7.69)	(4.19)	(4.47)	(0.56)
$D_{INDEXij,t}$	-0.00024	0.02035	0.00002	-0.00023	0.00003
	(-1.68)	(4.82)	(0.31)	(-1.81)	(1.48)
$D_{LISTING_{ij,t}}$	-0.00019	0.00310	0.00027	-0.00049	0.00004
	(-1.78)	(1.32)	(2.18)	(-4.16)	(2.09)

Table 13: **Connected Comovement: Additional Controls and Cross-sectional Variation**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, as well as the cross products of the return components, $(N_{CF,i,t+1}) * (-N_{DR,j,t+1})$ for the sample of stocks defined in Table 1. We estimate

$$\begin{aligned}\mathbf{y} \;=\; & a+\sum_{k=1}^{5}\sum_{l=1}^{5} b_{f-k,l}*F^{*}_{ij,t}+b_a*A^{*}_{ij,t}+b_s*DIFF\_SIZE^{*}_{ij,t}+b_b*DIFF\_BEME^{*}_{ij,t}\\ & +b_m*DIFF\_MOM^{*}_{ij,t}+b_k*NUM\_SIC^{*}_{ij,t}+b_{s1}*SIZE1^{*}_{ij,t}\\ & +b_{s2}*SIZE2^{*}_{ij,t}+b_{s12}*SIZE1SIZE2^{*}_{ij,t}+b_{ret}*RETCORR_{ij,t}\\ & +b_{roe}*ROECORR_{ij,t}++b_{vol}*VOLCORR_{ij,t}+b_{grth}*DIFFGRTH_{ij,t}\\ & +b_{lev}*DIFFLEV_{ij,t}+b_{state}*D_{STATE_{ij,t}}+b_{index}*D_{INDEXij,t}\\ & +b_{price}*DIFFPRICE_{ij,t}+b_{listing}*D_{LISTING_{ij,t}}+\varepsilon_{ij,t}\end{aligned}$$

where $\mathbf{y} = [r_{i,t+1}r_{j,t+1};\; N_{CF,i,t+1}N_{CF,j,t+1};\; -N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1};\; N_{DR,i,t+1}N_{DR,j,t+1}]$. The specification is similar to corresponding regressions in Tables 9 and 11, but we include in the regression (although they are not shown) additional variables as a robustness check. The additional variables are constructed as in Chen, Chen, Li (2009) and are as follows: past return correlation, $RETCORR_{ij,t}$; past profitability correlation, $ROECORR_{ij,t}$; the past correlation in the stocks abnormal trading volume, $VOLCORR_{ij,t}$, the absolute value of the difference in five-year log sales growth rates, $DIFFGRTH_{ij,t}$; the absolute difference in financial leverage ratios (defined as long-term debt / total assets), $DIFFLEV_{ij,t}$; the absolute value of the difference in the two stocks' log share prices, $DIFFPRICE_{ij,t}$; a dummy variable in the two firms are located in the same state; $D_{STATE_{ij,t}}$; a dummy variable if the two stocks belong to the same S&P major, mid-cap, or small-cap index, $D_{INDEXij,t}$; and a dummy variable if the two stocks are on the same stock exchange, $D_{LISTING_{ij,t}}$. All of these variables (except the dummies) are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. The return components are constructed from the aggregate and firm-level VARs estimated in Tables 2 and 3 as described in the Appendix. We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

PANEL A: Dependent var: $r_{i,t+1}r_{j,t+1}$						
Controls included but not shown						
$b_{f-k,l}$ estimates		Size of the pair (k)				
		Low	2	3	4	High
	Low	0.00160	0.00108	0.00083	0.00062	0.00052
		(3.17)	(3.89)	(4.74)	(5.56)	(5.40)
Total	2	0.00093	0.00086	0.00072	0.00056	0.00050
net		(4.50)	(5.35)	(6.09)	(5.95)	(5.30)
flow	3	0.00072	0.00068	0.00062	0.00051	0.00046
from		(5.67)	(6.55)	(6.51)	(6.14)	(5.44)
common	4	0.00097	0.00074	0.00064	0.00053	0.00045
funds		(6.15)	(7.17)	(8.02)	(6.98)	(5.82)
	High	0.00150	0.00127	0.00099	0.00077	0.00059
		(5.55)	(5.06)	(6.38)	(5.89)	(5.39)

PANEL B: Dependent var: $(N_{CF,i,t+1}) * (-N_{DR,j,t+1})$						
Controls included but not shown						
$b_{f-k,l}$ estimates		Size of the pair (k)				
		Low	2	3	4	High
	Low	0.00067	0.00052	0.00040	0.00031	0.00032
		(3.06)	(3.80)	(4.34)	(4.11)	(4.24)
Total	2	0.00041	0.00045	0.00038	0.00028	0.00033
net		(3.94)	(4.76)	(5.26)	(4.97)	(4.74)
flow	3	0.00034	0.00037	0.00035	0.00028	0.00031
from		(4.96)	(5.58)	(6.44)	(6.03)	(5.23)
common	4	0.00046	0.00040	0.00035	0.00028	0.00030
funds		(5.02)	(5.76)	(5.94)	(5.54)	(5.14)
	High	0.00059	0.00062	0.00050	0.00042	0.00042
		(3.90)	(4.32)	(5.04)	(4.60)	(4.90)

Table 14: **Connected Comovement**

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting measures of stock-pair comovement for the sample of stocks defined in Table 1. In particular, we forecast the realized cross-product of monthly returns, $r_{i,t+1}r_{j,t+1}$, the daily return Fisher correlation (ρ_{Fisher}), or $-N_{DR,j,t+1}N_{CF,i,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$ realized in month t+1. We estimate

$$\begin{aligned}\mathbf{y} \; = \; & a+\sum_{k=1}^{5}\sum_{l=1}^{5} b_{f-k,l}*F^{*}_{ij,t}+b_a*A^{*}_{ij,t}+b_s*DIFF\_SIZE^{*}_{ij,t}+b_b*DIFF\_BEME^{*}_{ij,t} \\ & +b_m * DIFF\_MOM^{*}_{ij,t} + b_{k}*NUM\_SIC^{*}_{ij,t}+b_{s1}*SIZE1^{*}_{ij,t} \\ & +b_{s2}*SIZE2^{*}_{ij,t}+b_{s12}*SIZE1SIZE2^{*}_{ij,t}+b_{ret} * RETCORR_{ij,t} \\ & +b_{roe} * ROECORR_{ij,t}+b_{\text{growth}} * DIFFGROWTH_{ij,t} + b_{state} * D_{STATE_{ij,t}} \\ & +b_{listing} * D_{LISTING_{ij,t}} + b_{index} * D_{INDEXij,t} + b_{price} * DIFFPRICE_{ij,t} \\ & +b_{lev}*DIFFLEV_{ij,t} + b_{vol} * VOLCORR_{ij,t} + \varepsilon_{ij,t}\end{aligned}$$

where $\mathbf{y}= [r_{i,t+1}r_{j,t+1};\rho_{Pearson}; -N_{DR,j,t+1}N_{CF,i,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}]$. The independent variables are updated quarterly and include our main measures of institutional connectedness, common funds ($F_{ij,t}$) and common analysts ($A_{ij,t}$), and a series of controls at time t. Each row varies the definition of common ownership for our benchmark specification (Panel A, as in Table 7) and our specification that includes the Chen, Chen, and Li variables (Panel B, as in Table 12). As measures of common ownership, we use the number of common owners, $F_{ij,t}$; the total ownership by all common funds in dollars of the two stocks scaled by the total market capitalization of the two stocks, $F^{CAP}_{ij,t}$; the average ownership stake by all common owners across the two stocks, $F^{SHARES}_{ij,t}$; and the total ownership by all common funds in dollars of the two stocks scaled by the Total Net Assets of all common owners, $F^{TNA}_{ij,t}$. All of these variables are then rank transformed and normalized to have unit standard deviation, which we denote with an asterisk superscript. We calculate Newey-West standard errors (four lags) of the Fama-MacBeth estimates that take into account autocorrelation in the cross-sectional slopes.

	Benchmark			All		
Variable	$r_{i,t+1}r_{j,t+1}$	ρ_{Fisher}	$-N_{DR,i}N_{CF,j}$ $-N_{DR,j}N_{CF,i}$	$r_{i,t+1}r_{j,t+1}$	ρ_{Fisher}	$-N_{DR,i}N_{CF,j}$ $-N_{DR,j}N_{CF,i}$
$F^{*}_{ij,t}$	0.00047	0.01952	0.00017	0.00050	0.01075	0.00027
	(6.36)	(13.95)	(3.94)	(6.43)	(11.77)	(5.61)
Avg R^2	0.82%	4.60%	1.09%	1.61%	6.40%	2.68%
$F^{CAP*}_{ij,t}$	0.00042	0.01056	0.00018	0.00036	0.00580	0.00020
	(6.83)	(13.70)	(6.31)	(6.48)	(7.06)	(5.69)
Avg R^2	0.79%	4.33%	1.04%	1.60%	6.38%	2.66%
$F^{SHARES*}_{ij,t}$	0.00029	0.00798	0.00018	0.00026	0.00569	0.00017
	(6.30)	(12.25)	(5.58)	(6.08)	(8.71)	(5.40)
Avg R^2	0.70%	4.25%	0.96%	1.53%	6.35%	2.58%
$F^{TNA*}_{ij,t}$	0.00044	0.01138	0.00014	0.00039	0.00516	0.00018
	(6.00)	(12.49)	(3.31)	(5.80)	(6.06)	(5.01)
Avg R^2	0.79%	4.34%	1.07%	1.59%	6.36%	2.65%

Table 15: **Alphas on Connected Trading Strategies**

This table presents the profitability of a simple trading strategy exploiting stock connectedness. We independently sort stocks into quintiles based on their own return over the last three months and the return on their connected portfolio over the last three months. We measure the connected return as $r_{iC,t} = \sum_{j=1}^{J} F^{**}_{ij,t-1} r_{j,t} / \sum_{j=1}^{J} F^{**}_{ij,t-1}$ where $F^{**}_{ij,t} = F^{*}_{ij,t}$ if $F_{ij,t} > 0$ and $F^{**}_{ij,t} = 0$ if $F_{ij,t} = 0$. Each portfolio holds the associated stocks for the next five months. We estimate coefficients from monthly regressions of

$$r_{p,t} - r_{f,t} = \alpha_5 + bRMRF_t + sSMB_t + hHML_t + mMOM_t + rSTREV_t + \varepsilon_{p,t}$$

where $r_{p,t}$ is the equal-weight excess return on the portfolio of the stocks associated with the particular trading strategy. Panel A reports alphas where the factor STREV is excluded from the regression and Panel B reports α_5.

PANEL A: FOUR FACTOR ALPHAS

		Connected portfolio					
		Low	2	3	4	High	L - H
Own return	Low	0.0047	0.0045	0.0032	0.0018	0.0017	0.0029
		(3.53)	(4.18)	(2.93)	(1.55)	(1.42)	(1.79)
	2	0.0045	0.0037	0.0025	0.0021	0.0008	0.0038
		(3.96)	(4.07)	(2.81)	(2.33)	(0.74)	(2.74)
	3	0.0030	0.0017	0.0009	-0.0000	0.0001	0.0029
		(2.77)	(1.94)	(1.10)	(-.04)	(0.10)	(2.24)
	4	0.0023	-0.0002	-0.0011	-0.0012	-0.0013	0.0036
		(1.91)	(-.26)	(-1.4)	(-1.5)	(-1.6)	(2.73)
	High	-0.0001	-0.0003	-0.0025	-0.0023	-0.0018	0.0017
		(-.11)	(-.24)	(-3.0)	(-2.7)	(-1.8)	(1.04)
	L - H	0.0048	0.0047	0.0057	0.0042	0.0035	0.0065
		(3.52)	(4.03)	(4.46)	(3.27)	(2.56)	(3.74)

PANEL B: FIVE FACTOR ALPHAS

		Connected portfolio					
		Low	2	3	4	High	L - H
Own return	Low	0.0044	0.0040	0.0028	0.0014	0.0015	0.0028
		(3.27)	(3.74)	(2.54)	(1.16)	(1.23)	(1.72)
	2	0.0044	0.0035	0.0022	0.0019	0.0005	0.0039
		(3.84)	(3.81)	(2.55)	(2.10)	(0.47)	(2.85)
	3	0.0028	0.0015	0.0007	-0.0003	-0.0002	0.0030
		(2.52)	(1.71)	(0.81)	(-.35)	(-.20)	(2.24)
	4	0.0021	-0.0006	-0.0014	-0.0013	-0.0015	0.0036
		(1.71)	(-.62)	(-1.8)	(-1.7)	(-1.8)	(2.66)
	High	-0.0006	-0.0006	-0.0028	-0.0025	-0.0019	0.0013
		(-.46)	(-.60)	(-3.3)	(-2.9)	(-1.8)	(0.78)
	L - H	0.0050	0.0046	0.0056	0.0038	0.0034	0.0062
		(3.59)	(3.89)	(4.32)	(2.99)	(2.42)	(3.56)

Table 16: **The Connected Strategy and Liquidity Risk**

This table measures the loadings of the connected sstock trading trategy on two common liquidity factors as well as on time effects. We regress the returns of the connected strategy on a constant, liquidity factors from the work of Pastor and Stambaugh (2003), PS_INNOV, and Sadka (2006), $SADKA\_PV$, the Fama-French/Carhart factors, a short-term reversal factor, a trend, and seasonal (quarterly) dummies. Columns 1 and 2 report loadings of our connected strategy on both liquidity factors for the period March 1983 to December 2005 (Sadka's liquidity factor is only available during that period). Columns 3 to 5 show the loadings of the connected strategy to the PS liquidity factor, a trend, and quarterly seasonal dummies, for the period June 1980 to December 2008.

	Dependent Variable: Connected Strategy			
	1	2	3	4
Alpha	0.0063	0.0066	0.0065	0.0112
	(3.24)	(3.35)	(3.79)	(3.42)
PS_INNOV	0.0633		0.0487	
	(1.94)		(1.71)	
SADKA_PV		0.3919		
		(1.16)		
Mktrf	0.0146	0.0393	-0.0164	0.0116
	(0.28)	(0.78)	(-0.36)	(0.26)
SMB	-0.4058	-0.4046	-0.3637	-0.3437
	(-6.78)	(-6.71)	(-6.59)	(-6.13)
HML	-0.1252	-0.1191	-0.1814	-0.1523
	(-1.73)	(-1.64)	(-2.83)	(-2.34)
UMD	-0.9038	-0.9189	-0.9102	-0.9147
	(-20.52)	(-20.38)	(-22.25)	(-22.31)
ST_Reversal	0.0277	0.0208	0.0070	0.0075
	(0.48)	(0.36)	(0.13)	(0.14)
Trend			0.0000	
			(-0.63)	
Q1				-0.0070
				(-1.48)
Q2				-0.0086
				(-1.87)
Q3				-0.0032
				(-0.70)
Obs	274	274	343	343
Rsquare	0.68	0.68	0.66	0.66

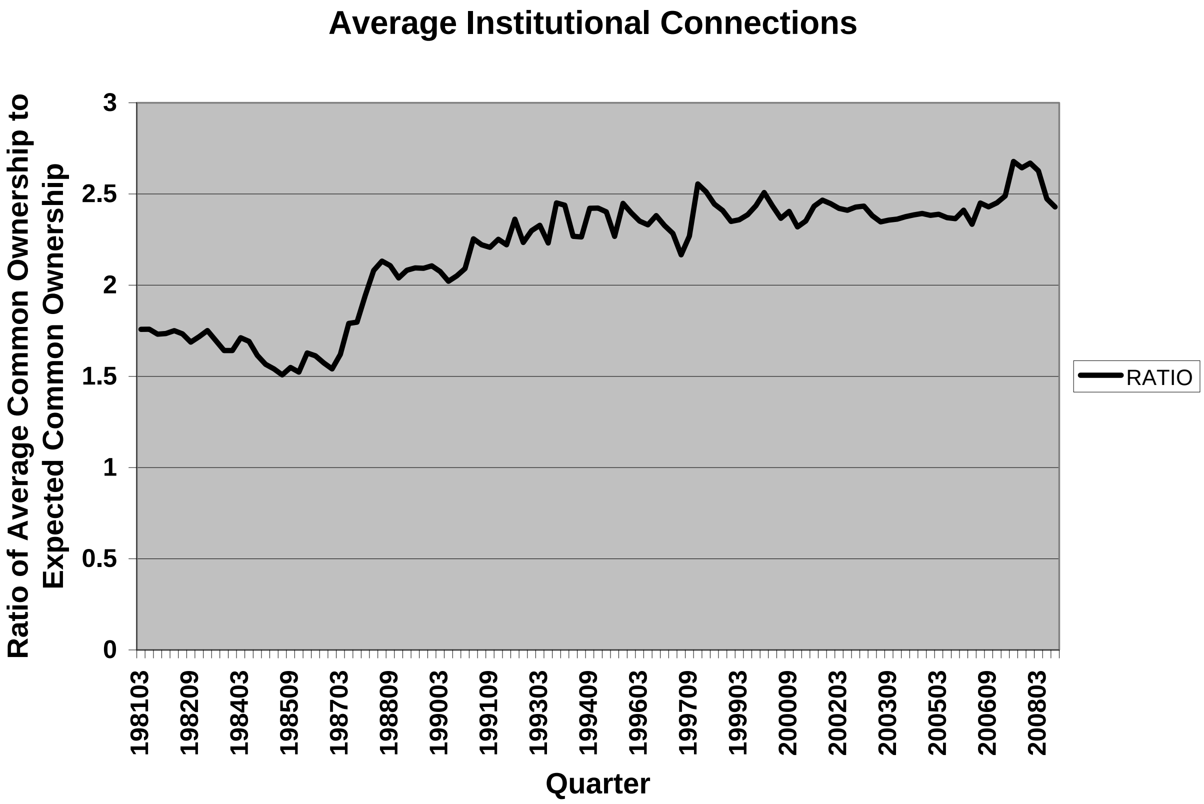

Figure 1: This figure plots the time-series evolution of the ratio of the average number of common funds per pair in each cross section of stock pairs to the average number of common funds per pair if all funds in that cross section held the same number of stocks as the average fund holds.

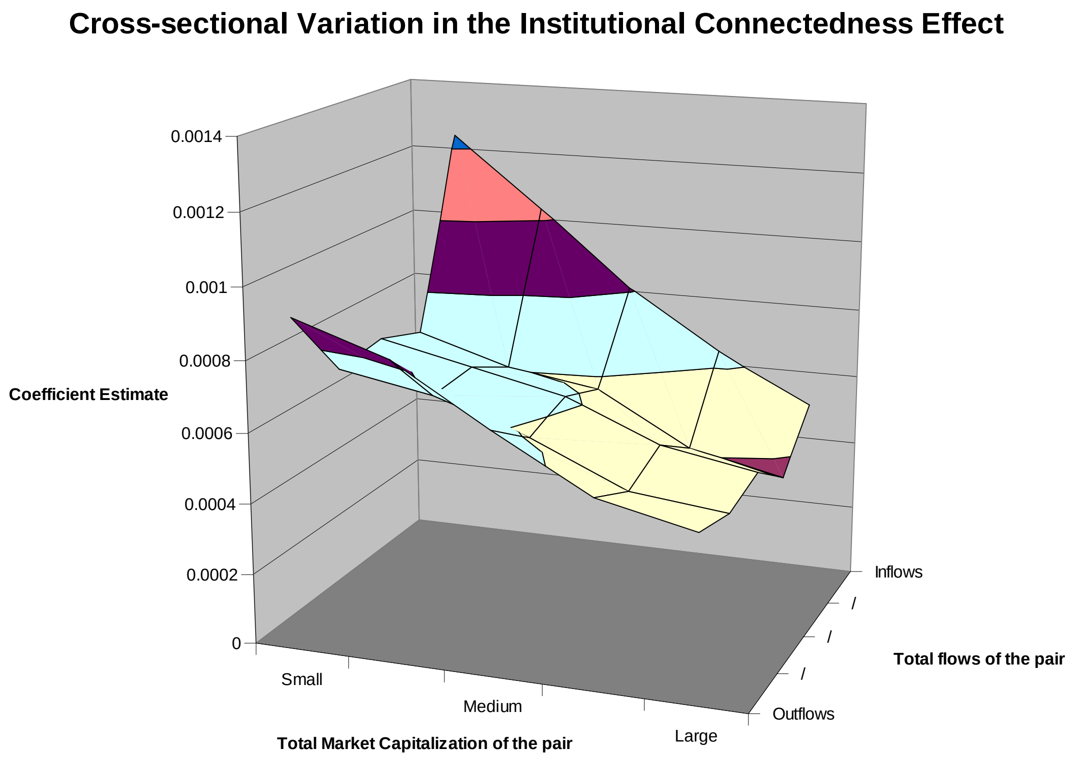

Figure 2: This figure plots the point estimates from Table 9. In that table we interact the coefficient on the number of common funds per pair with dummies for the size of the pair of stocks and the total net flow into the common funds. Specifically, each year we sort pairs into quintiles based on their total market capitalization. We independently sort pairs into quintiles based on their total net flow. Thus the interactions reflect the cross-sectional variation in stock-pair heterogeneity.

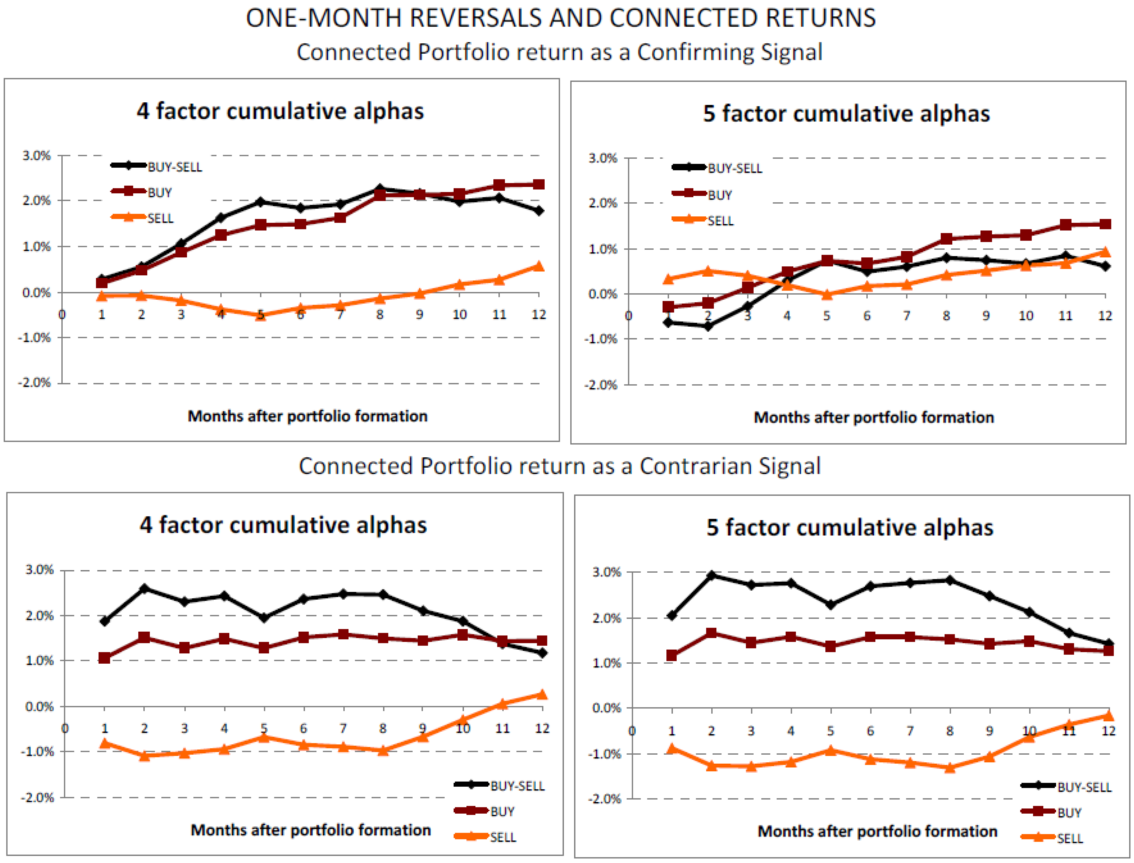

Figure 3: This figure graphs the abnormal performance of buy-and-hold strategies that trade the one-month reversal strategy conditional on the return on a stock's connected portfolio. Stocks are sorted into 25 portfolios based on independent quintile sorts on a stock's own one-month return and its one-month connected return. The top half of the figure buys (sells) stocks whose own returns are relatively low (high and whose connected returns are relatively low (high). The bottom half of the figure buys (sells) stocks whose own returns are relatively low (high) and whose connected returns are relatively high (low). The left side of the figure benchmarks returns against the Carhart four-factor model while the right side of the figure benchmarks returns against the Carhart model augmented with the one-month reversal factor.

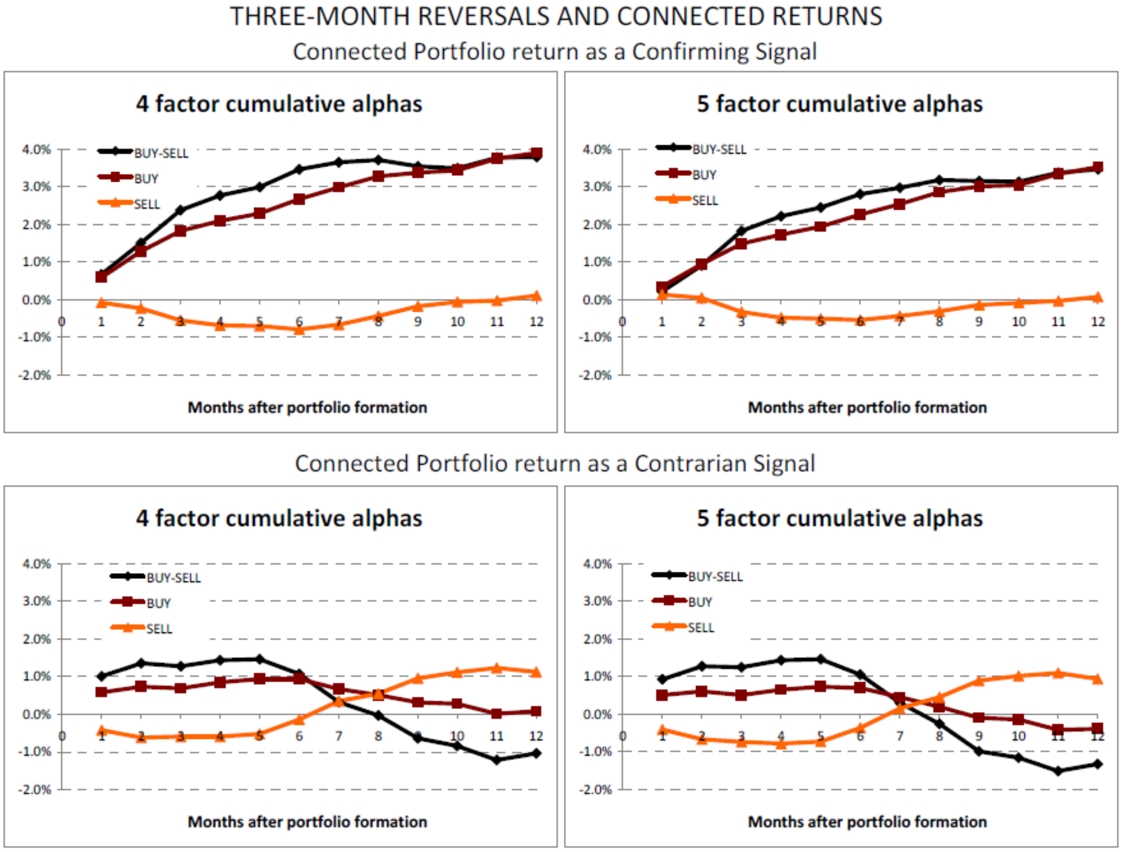

Figure 4: This figure graphs the abnormal performance of buy-and-hold strategies that trade a three-month reversal strategy conditional on the return on a stock's connected portfolio. Stocks are sorted into 25 portfolios based on independent quintile sorts on a stock's own three-month return and its three-month connected return. The top half of the figure buys (sells) stocks whose own returns are relatively low (high and whose connected returns are relatively low (high). The bottom half of the figure buys (sells) stocks whose own returns are relatively low (high) and whose connected returns are relatively high (low). The left side of the figure benchmarks returns against the Carhart four-factor model while the right side of the graphs benchmarks returns against the Carhart model augmented with the one-month reversal factor.