THREE ESSAYS ON THE COMOVEMENT OF FINANCIAL ASSETS

Miguel Antón Sancho

Dissertation submitted for the degree of

Doctor of Philosophy in Finance at

The London School of Economics and Political Science

July 2011

Declaration

I certify that the thesis I have presented for examination for the PhD degree of the London School of Economics and Political Science is solely my own work other than where I have clearly indicated that it is joint work.

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Miguel Antón Sancho

Abstract

In this thesis I study the effects of institutional trading on the comovement of financial assets. In the first chapter, joint work with Christopher Polk, we connect stocks through common active mutual fund ownership, and use these connections to forecast cross-sectional variation in return covariance, controlling for similarity in style and other pair characteristics. We argue this covariance is due to contagion based on return decomposition evidence, cross-sectional heterogeneity in the extent of the effect, and the magnitude of average abnormal returns to a cross-stock reversal trading strategy exploiting information in these connections. We show that the typical long/short hedge fund covaries negatively with this strategy suggesting that hedge funds may potentially exacerbate the price dislocation we document. In the second chapter I study the sources of change in the systematic risks of stocks added to the S&P 500 index. Firstly, using vector autoregressions (VARs) and a two-beta decomposition, I find that I cannot reject the hypothesis that all of the well-known change in beta comes from the cash-flow news component of a firm's return. Secondly, I study fundamentals of included firms directly to reduce any concerns that the VAR-based results are sensitive to my particular specification. As ownership structure cannot directly influence fundamentals, these results challenge previous findings, as they are consistent with the change in beta being due to a selection effect. In the third chapter, joint work with Daniel Bergstresser, we explore index-based comovement in the market for Credit Default Swaps (CDS). We exploit the additions of individual CDS contracts in the Markit CDX Index, a major credit derivative benchmark. We find that for single name CDS contracts, comovement increases after inclusion in the index. Comparing movements in the CDS spreads to movements of the bonds of the same issuers, the CDS spread comovement increases significantly more than the bond spread comovement. This pattern of evidence is consistent with the excess comovement in equity markets documented by Barberis et al (2005) and others.

Acknowledgements

First and foremost I would like to thank my advisor, Christopher Polk. I am very grateful for his constant support and guidance, and for the opportunities he has provided me with. Special thanks goes too to my second advisor, Dimitri Vayanos. Their availability, encouragement, and continued support have been invaluable. I would also like to thank all the faculty members of the Department of Finance at the LSE, for their helpful discussions and feedback in seminar presentations. Thanks to Andrew Patton as well, my very first advisor at LSE, for his initial support.

I also thank all my friends and fellow researchers at the LSE, and FMG in particular, for discussions and comments on my work. Special mention goes to Vincent and Fabián, with whom I spent so many hours, in the good times... and in the bad ones. I will keep so many good memories of the job market with my fellow classmates, Vincent, Daniel, Dragana, Yuki, Gyuri, Aytek, Siti, and Zhigang. Good mention goes as well to Brando and Castañeda in a particular way. I would like to thank Mary and Osmana for their always being there. A very particular thanks goes to all my students at the LSE, I really enjoyed teaching them, I wish the best for all of them. Of course, special thanks to Antoine, for opening the doors of LSE to me. It is too bad that you left us two years ago now. I am sure you keep taking care of your students from above.

Thanks are also due to John Campbell for facilitating my visit to the Economics Department at Harvard University during the academic year 2008-2009. I truly enjoyed my time in Boston and learned a lot from the courses and seminars at Harvard. Special thanks go to Daniel Bergstresser for his good humor and his help, it has been great working with him. I also thank Luis Viceira, Robin Greenwood, Borja Larrain, Josh Coval, Lauren Cohen, and Chris Malloy for their time and helpful discussions. I made good friends while at Harvard. I spent many hours in the squash courts with Daniel Sunesson, and in Vermont with Juanjo. It was very nice to work closely at HBS with Johnny Kang, Sergey Chernenko, and Bob Turley. Pato, Johnny, and Dwight made my stay in Boston very pleasant.

Financial support from the Fundación Ramón Areces, and from the LSE is gratefully acknowledged.

Luis Ravina and Antonio Moreno have been a great support ever since my undergraduate degree. I owe a lot to my friends for their support over the past few years. Alvaro, Peter, Jorge, Pablo, Joe, Quique, Sergio, Francis, Javi Muñoz, Ricky, Carlo, Arnil, Miguel, Marc, Alex, Prakarsh, and Aymeric deserve special mention. I have also benefited enormously from the great atmosphere of Netherhall House.

I dedicate this thesis to my parents, to my siblings María, Javi, Leticia, Pablo, Alvaro, Maialen, and my aunt Margarita. They make it all worthwhile.

Contents

List of Figures

List of Tables

1 Connected Stocks

(Joint work with Christopher Polk)

1.1 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 may layer on additional covariation as well. 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. This finding continues to hold after controlling for a wide variety of other pair characteristics in addition to these standard style controls.

We provide evidence consistent with common ownership causing the increased covariation associated with ownership. First, 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, in strong contrast to the ownership results. Second, 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.

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. In particular, 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 up to 7% per year, controlling for market, size, value, momentum, 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 the typical hedge fund and in particular the typical long-short hedge fund load negatively on our trading strategy. In fact, the exposures of these value-weight hedge fund indexes are more negative than the corresponding exposure of a value-weight portfolio of the active mutual funds in our sample. This suggests that the typical hedge fund may be part of the problem (creating the covariance) instead of part of the solution.[1]

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

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

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 (2010) 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.[2] In particular, 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.[3] 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 institutional-driven comovement follows naturally from that theory and empirical evidence.

Four recent working papers analyze issues related to stock return comovement and/or institutional 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 persistence, 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

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

[3]Many researchers have built on the ideas in Shleifer and Vishny (1997), including Gromb and Vayanos (2002), Vayanos (2004), and Brunnermeier and Pedersen (2009). For a recent survey of this literature, see Gromb and Vayanos (2010).

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 covaries 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 distribution 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 pair-wise correlations and show that a large portion of that cross-sectional variation is persistent, 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) explore whether hedge funds take advantage of the mutual fund flow-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. Our findings are consistent with their results but further show that the typical hedge fund apparently winds up 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.

1.2 Methodology

1.2.1 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 $A^*_{ij,t}$.

1.2.2 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 the 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, NUM_SIC.

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 greater similarity in the characteristics mentioned above. To measure this similarity, each quarter we first calculate every stock's percentile ranking on a particular firm characteristic. Our measures of similarity, $SAME\_SIZE$, $SAME\_BEME$, and $SAME\_MOM$, are then just the negative of the absolute difference in percentile ranking across a pair for a particular characteristic. As with our institutional connectedness measures, we do not use these variables directly but instead work with normalized rank transforms, which we continue to denote with an asterisk superscript. 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:

$$\begin{aligned} r_{i,t+1}r_{j,t+1} \;=\;& a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} + b_s * SAME\_SIZE^*_{ij,t} \\ & + b_b * SAME\_BEME^*_{ij,t} + b_m * SAME\_MOM^*_{ij,t} \\ & + b_k * NUM\_SIC^*_{ij,t} + b_{s1} * SIZE1_{ij,t} + b_{s2} * SIZE2_{ij,t} \\ & + b_{s12} * SIZE1SIZE2^*_{ij,t} + \varepsilon_{ij,t}. \end{aligned} \tag{1.1}$$

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:

$$
\begin{aligned}
r_{i,t+1}r_{j,t+1} &= a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} \\
&+ \sum_{s=0}^{9} b_s * D_{DIFF\_SIZE_{ij,t}=s} + \sum_{b=0}^{9} b_b * D_{DIFF\_BEME_{ij,t}=b} \\
&+ \sum_{m=0}^{9} b_m * D_{DIFF\_MOM_{ij,t}=m} + \sum_{k=0}^{3} b_k * D_{NUM\_SIC_{ij,t}=k} \\
&+ b_{s1} * SIZE1_{ij,t} + b_{s2} * SIZE2_{ij,t} \\
&+ b_{s12} * SIZE1SIZE2^*_{ij,t} + \varepsilon_{ij,t}
\end{aligned} \tag{1.2}
$$

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 as well as normalized to have unit standard deviation so that the intercept a measures the average cross-sectional effect and the regression coefficients are easily interpreted. 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.

1.2.3 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 the CDA / Spectrum database provided by Thomson Reuters and the CRSP Mutual Fund database. We use the Mutual Fund Links dataset created by Russ Wermers 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.[4] In addition, for a fund to be in our sample we require it to hold at least one stock in our stock sample at a point in time.

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 1.1 confirms the well-known marked increase in funds over this period. The number of analysts has also increased, though not as dramatically. Table 1.2 reports 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 1.3 reports various summary statistics for returns and the news components. Consistent with Vuolteenaho (2002), cash-flow news makes up a larger portion of total return variance.

1.3 Results

Table 1.3 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 nine. In the later part of the sample (2000-2007), that median number increased to

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

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 Table 1.4. 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 1.4 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 1.1 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 1.5 Panel A reports the result of our forecasting cross-sectional variation in realized cross products. We begin by estimating simpler versions of equation (1.1). 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% of the average amount of covariation. In column (2) of Table 1.5 Panel A, 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 1.5 Panel A 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 similarity in this style 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), but 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) of Table 1.5 Panel A, 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 the extent of institutional ownership of a stock. 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, with the coefficient on common ownership doubling in magnitude.[5]

The final column of Table 1.5 Panel A generalizes our controls for stock similarity by turning to dummy variables to capture the difference in size, beme, or momentum decile across the pair.[6] We also dummy the number of common SIC digits. We report these dummy coefficient estimates of equation (1.2) in Panel B of Table 1.5. 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 1.6, we use alternative measures of comovement between two stocks. In the first column of Table 1.6, we repeat the estimates from the fourth column of Table 1.5 Panel A (our full benchmark specification) for ease of comparison. In column (2), we keep the same control variables as in the full benchmark specification of Table 1.5 Panel A 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 1.5 but replace the monthly return cross product with Pearson and Fisher 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 1.6 ease concerns of our use of the realized monthly return cross product (and its

[5] Note that by including these additional size controls, the coefficient on $SAME\_SIZE^*_{ij,t}$ changes sign due to the correlation among the size variables.

[6] Note that our dummies are for the difference in characteristic deciles across the firms in a pair, so that one's prior of the sign of the coefficient should be the negative of that in Panel A of the Table.

components) throughout the rest of the paper.

To summarize, the main conclusion from Tables 1.5 and 1.6 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 hope that our method and model will perform relatively well out-of-sample, consistent with the related claims of Brandt, Santa-Clara, and Valkanov (2009).

1.3.1 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 five-year monthly 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), $DIFFLEV_{ij,t}$; 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 S&P 500 index, $D_{INDEXij,t}$; and a dummy variable if the two stocks are on the same stock exchange, $D_{LISTING_{ij,t}}$. Thus our specification is

$$
\begin{aligned}
\mathbf{y}_{ij,t+1} &= \qquad (1.3) \\
& [r_{i,t+1} r_{j,t+1}; \rho_{ij,Fisher}] \\
&= a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} + b_s * SAME\_SIZE^*_{ij,t} \\
& + b_b * SAME\_BEME^*_{ij,t} + b_m * SAME\_MOM^*_{ij,t} \\
& + b_k * NUM\_SIC^*_{ij,t} + b_{s1} * SIZE1^*_{ij,t} + b_{s2} * SIZE2^*_{ij,t} \\
& + b_{s12} * SIZE1SIZE2^*_{ij,t} \\
& + b_{ret} * RETCORR^*_{ij,t} + b_{roe} * ROECORR^*_{ij,t} \\
& + b_{vol} * VOLCORR^*_{ij,t} + b_{grth} * DIFFGRTH^*_{ij,t} \\
& + b_{lev} * DIFFLEV^*_{ij,t} + b_{price} * DIFFPRICE^*_{ij,t} \\
& + b_{state} * D_{STATE_{ij,t}} + b_{index} * D_{INDEXij,t} \\
& + b_{listing} * D_{LISTING_{ij,t}} + \varepsilon_{ij,t}
\end{aligned}
$$

where $\mathbf{y}$ is either the realized cross product or the realized Fisher return correlation over the next month.

The first two regressions in Table 1.7 repeat the key regressions from Tables 1.5 and 1.6, but including these additional controls. In particular, in regression 2 of Table 1.7, we reproduce the essence of 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 also 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). Finally, 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. We return to the three remaining columns of Table 1.7 in the next section.

Table 1.8 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 net assets of all common owners across the two stocks, $F_{ij,t}^{TNA}$. Our next alternative is to measure common ownership as the total dollar ownership by all common funds of the two stocks scaled by the total market capitalization of the two stocks, $F_{ij,t}^{\%CAP}$. Finally, we use as our last measure the total dollar ownership by all common funds of the two stocks scaled by the Total Net Assets of all common owners, $F_{ij,t}^{\%TNA}$. In this section, we focus on the first two columns of each Panel. All definitions continue to forecast cross-sectional variation in the realized return cross-product (the first regression in each Panel) and the subsequent return correlation (the second regression in each Panel). 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. We return to the third column of each Panel in Table 1.8 in the next section.

1.3.2 Connectedness and temporary components of returns

Tables 1.5, 1.6, 1.7, and 1.8 document 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 as in Coval and Stafford (2007) and Lou (2009). To provide additional evidence that this is the case, we first 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 move 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 methodology we now follow is very similar to the one described above, but we change the left hand side of equation 1.3. Specifically, the new equation we estimate has the form:

$$\begin{aligned}
\mathbf{y}_{ij,t+1} &= \\
& [N_{i,CF_{t+1}} N_{j,CF_{t+1}}; -N_{i,CF_{t+1}} N_{j,DR_{t+1}} - N_{j,CF_{t+1}} N_{i,DR_{t+1}}; \\
& N_{i,DR_{t+1}} N_{j,DR_{t+1}}] \\
&= a + b_f * F^*_{ij,t} + b_a * A^*_{ij,t} + b_s * SAME_SIZE^*_{ij,t} \\
& + b_b * SAME_BEME^*_{ij,t} + b_m * SAME_MOM^*_{ij,t} \\
& + b_k * NUM_SIC^*_{ij,t} + b_{s1} * SIZE1^*_{ij,t} + b_{s2} * SIZE2^*_{ij,t} \\
& + b_{s12} * SIZE1SIZE2^*_{ij,t} \\
& + b_{ret} * RETCORR^*_{ij,t} + b_{roe} * ROECORR^*_{ij,t} \\
& + b_{vol} * VOLCORR^*_{ij,t} + b_{grth} * DIFFGRTH^*_{ij,t} \\
& + b_{lev} * DIFFLEV^*_{ij,t} + b_{price} * DIFFPRICE^*_{ij,t} \\
& + b_{state} * D_{STATE_{ij,t}} + b_{index} * D_{INDEXij,t} \\
& + b_{listing} * D_{LISTING_{ij,t}} + \varepsilon_{ij,t}
\end{aligned} \tag{1.4}$$

where $\mathbf{y}$ is a vector of the various components of the realized return cross-product. The results of our covariance decomposition can be found in the third, fourth, and fifth regressions of Table 1.7. In the third regression, we find that a modest but statistically 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 0.00010. As argued above, this component must reflect the choices that fund owners make. The fifth regression in Table 7 shows that there is also a statistically significant but even less economically important 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.

The fourth column of Table 1.7 reports the main finding of this section. 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 0.00027, with a t-statistic of 5.59. Note that the average effect is -.00076. 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 ownership, return covariation is increased.

The finding that the typical interaction between cash-flow news and discount-rate news across stocks reduces covariance is consistent with the findings of Vuolteenaho (2002), who finds that the typical stock's cash-flow news is positively correlated with its own discount-rate news, reducing firm volatility. Vuolteenaho interprets this finding as being consistent with a simple story where the typical project is zero NPV. Given his results, it comes as no surprise that the typical cross-stock effect is negative. In this context, our finding that the ownership-based component increases covariance is all the more striking.

Interestingly, the ability of common analyst coverage to forecast subsequent return covariation mainly arises from the covariation of cash-flow news of one stock with the cash-flow news of another. The fact that the common coverage institutional connection works differently than the common ownership institutional connection makes the price pressure interpretation of the main finding of this section more compelling.

The third regression in each Panel of Table 1.8 investigates the impact of varying the definition of our measure of institutional connectedness on the ability of common ownership to forecast this component of the return covariance. All four measures appear to be capturing the component of return covariance that is due to the covariance of the discount-rate news of one stock in the pair with the cash-flow news of the other stock in the pair.

1.3.3 When does connectedness matter?

To provide additional evidence in support of the causal interpretation, we now exploit cross-sectional heterogeneity in stock pair characteristics. Specifically, in Table 1.9, 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:

$$FLOW_{i,t} = \frac{TNA_{i,t} - TNA_{i,t-1} * (1 + R_{i,t})}{TNA_{i,t-1}} \tag{1.5}$$

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 monthly flows, then we sum them up and finally we divide them by the previous quarter TNA.

Panel A of Table 1.9 estimates the interaction for the benchmark specification of Table 1.5. 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 **??** shows these patterns graphically.

In Panel B of Table 1.9, we repeat our exercise of interacting 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 for the full specification of Table 1.7. Consistent with our interpretation, Panel B of Table 1.9 shows that the cross-sectional variation in the magnitude of the coefficient documented in Table 1.9 Panel A also shows up in the full specification.

1.3.4 Connected trading strategies

Here we measure the profits to various trading strategies 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 quintiles 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 first consider two simple 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. We interpret such a strategy as exploiting the price pressure induced by common ownership. 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. This quite different bet would be consistent with a standard pairs trading strategy or with industry momentum. 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

$$r_{p,t+1} - r_{f,t+1} = \alpha_5 + bRMRF_{t+1} + sSMB_{t+1} + hHML_{t+1} + mMOM_{t+1} + rSTREV_{t+1} + \varepsilon_{p,t+1} \quad (1.6)$$

where the factors are the four factors of Fama and French (1993) and Carhart (1997), augmented with the short-term reversal factor.[7] 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 1.3 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 buy-and-hold 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 1.4 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 and the previous three-month return on the connected portfolio. The cumulative abnormal 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 evaluate the average returns on portfolio sorts that take these predictable patterns in the cross section of average returns into account. Table 1.10 reports the four and five-factor alphas from independent portfolio sorts based on the past three-month return on the own stock and the past three-month connected portfolio return. To further ensure our strategies do not merely reflect the standard one-month reversal effect, we first skip a month after the sort and then hold the stocks in question for five months, following the methodology of Jegadeesh and Titman (1993).

There are two general patterns in Table 1.10 that are consistent with our

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

initial conclusions concerning Figures 1.3 and 1.4. 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. As a consequence, we design two composite connected stocks trading strategies that use the connected return as a confirming signal.

The first strategy, $CS1$, buys the low own return / low connected return portfolio and sells the high own return / high connected return portfolio so that its return is $r_{CS1} = r_{low\ own\ /\ low\ connected} - r_{high\ own\ /\ high\ connected}$. The five-factor alpha for $CS1$ is an impressive 57 basis points per month with a corresponding t-statistic of 2.95. The second strategy, $CS2$, buys the average (across the own return quintiles) low connected return portfolios and sells the average (across the own return quintiles) high connected return portfolios so that its return is $r_{CS2} = \overline{r_{low\ connected}} - \overline{r_{high\ connected}}$. This strategy earns 32 basis points per month, with a t-statistic of 2.60. Though this strategy ignores the information in the interaction between a stock's own return and its connected return, the performance is still strong. For completeness, we plot the corresponding cumulative abnormal buy-and-hold performance of this strategy in Figure 1.5.

Table 1.11 includes additional explanatory variables, in particular a linear time trend, and end-of-quarter dummies, in the performance attribution of our first connected stocks trading strategy, $CS1$. We also include the liquidity factors of Sadka (2006) and Pastor and Stambaugh (2003). Documenting that our connected stocks trading strategy covaries with these popular measures of liquidity provides further confirmation of the source of the abnormal return. We do find that $CS1$ positively covaries with the non-traded liquidity factor of Pastor and Stambaugh across all of the specifications we consider.[8] Our $CS1$ strategy also covaries with the Sadka factor, though the result is not statistically significant. Similar conclusions hold for a version of Table 1.11 (not shown) analyzing the second connected stocks trading strategy, $CS2$.

[8]We has also used the traded factor of Pastor and Stambaugh (2003) as a sixth factor in our performance attribution. The abnormal returns on our connected strategy remain economically and statistically significant

1.3.5 Hedge Fund Index attribution

Our last analysis uses our two connected stocks trading strategies, $CS1$ and $CS2$, 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 one 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 in this investing space, this index gives a good representation of the extent to which our connected stock strategy reflects the general health of the hedge fund industry.[9] We also examine the performance of the long/short component of the CSFB index to measure the extent to which funds that specifically invest in equities are exposed to the connected stocks factor.

Table 1.12 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 strategy. The coefficient in the first column of Panel A in Table 1.12 estimates a regression of the overall hedge fund index excess return on the return on our first connected strategy, r_{CS1}, and the four factors of Fama and French (1993) and Carhart (1997), augmented with the short-term reversal factor. The coefficient is -0.0658 with a t-statistic of -2.08. 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 (2001, 2004).[10] 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 is now more economically and statistically significant; the point estimate is now -0.1114 and has an associated t-statistic of -6.09. Both results suggest that our trading strategy is useful tool to measure the state of the hedge fund industry.

Perhaps more interesting results are in columns 3 and 4 of Table 1.12. In column 3, we measure the degree to which the Long/Short subset of hedge funds

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

[10] We downloaded three of the Fung and Hsieh (2001) factors from http://faculty.fuqua.duke.edu/~dah7/DataLibrary/TF-FAC.xls.

covaries with our connected return trading strategy in the presence of the Fama-French/Carhart factors and the short-term reversal factor. In column 4, we use the Fung and Hsieh factors as controls instead. In both cases, we find that the returns on this subset of hedge funds strongly negatively covary with our connected return factor with loadings that are approximately 25-50% larger in absolute value. The t-statistics are correspondingly larger. This finding is very comforting as one would expect this subset of hedge funds to be more exposed to our factor.

For the sake of 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 -0.0265 with an associated t-statistic of -2.65. Though we do not observe complete holdings data for all hedge funds and therefore cannot see the exact positions of these long/short hedge funds, these results suggest that these hedge funds do not take full advantage of the opportunities that price pressure from mutual fund flows provide. In fact, one can argue that perhaps hedge funds are exacerbating rather than mitigating the price pressure patterns documented in this paper. Panel B of Table 1.12 repeats the analysis replacing r_{CS1} with r_{CS2}, the version of our connected strategy that ignores the information in the interaction between a stock's own return and its connected return. We find results that are qualitatively similar. In particular, the loading on r_{CS2} is statistically and economically significant. Additionally, the loading for the Long/Short subset of hedge funds is again much larger in absolute magnitude.

Figure 1.6 provides evidence on why it is not surprising that the typical hedge fund loads negatively on our connected strategy. This figure plots both the loadings of the two hedge fund indexes on the connected strategies as well as the cumulative abnormal return on the connected strategy in event time, where the event is the forming of the connected stock trading strategy (either $CS1$ or $CS2$). One reasonable interpretation of this figure is that hedge funds follow a momentum strategy that effectively front-runs mutual funds in distress. However, the typical hedge fund is unable to exit its positions in time and therefore exacerbates the price dislocation they help initiate.

1.4 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 up to 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 may be part of the cause of the excess covariation and price dislocation that contagion from ownership-based connections generates.

1.5 Tables

Table 1.1: 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. We show only the statistics for the last quarter of each year in our sample. 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 1.2: Aggregate and Firm-level VAR

Panel A 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. Panel B 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.

PANEL A: Aggregate VAR

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)	

PANEL B: Firm-level VAR

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 1.3: Ownership, Coverage, and Stock Returns: Summary Statistics

This table reports summary statistics for the sample defined in Table 1.1 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 41,374,135 pair-quarters. 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 1.4: Distribution of Common Fund Ownership and Analyst Coverage

Panel A reports the distribution of the variable $F_{ij,t}$ measuring the number of funds holding both stocks in a pair over the last quarter. Panel B 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. The distribution is shown for the average of all the sample (ALL), for the first and the last year in the sample (1983 and 2007 respectively), and for every five years. There are 41,374,135 pair-quarters.

PANEL A: The Cross-sectional Distribution of Common Fund Ownership

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
1985	0.89	1.77	0	0	0	1	4	8	58
1990	2.87	4.63	0	0	1	4	11	21	115
1995	8.14	10.38	0	2	5	11	26	49	231
2000	14.86	21.89	0	4	8	19	47	106	543
2005	22.80	24.35	0	8	15	29	64	120	500
2007	25.73	23.51	0	12	19	32	66	121	463

PANEL B: The Cross-sectional Distribution of Common Analyst Coverage

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
1985	0.42	1.86	0	0	0	0	2	9	48
1990	0.39	1.97	0	0	0	0	1	10	53
1995	0.25	1.41	0	0	0	0	1	7	39
2000	0.16	1.07	0	0	0	0	1	4	40
2005	0.16	1.20	0	0	0	0	0	5	43
2007	0.16	1.18	0	0	0	0	0	5	37

Table 1.5: 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 1.1. 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 measure the negative of the absolute value of the difference in size, BE/ME and momentum percentile ranking across the two stocks in the pair ($SAME\_SIZE_{ij,t}$, $SAME\_BEME_{ij,t}$, and $SAME\_MOM_{ij,t}$ 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 percentile rankings as well as the similarity in SIC code across the pair with a full set of dummy variables, which we report in Panel B. (Note that the dummy variables in Panel B now capture the difference in style across the pair, as described in the text.) 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)
$SAME\_SIZE^*_{ij,t}$			0.00002	-0.00028	
			(1.17)	(-4.77)	
$SAME\_BEME^*_{ij,t}$			0.00012	0.00009	
			(2.78)	(2.30)	
$SAME\_MOM^*_{ij,t}$			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				
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 1.6: 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 1.1. In particular, we forecast the realized cross-product of monthly returns, $r_{i,t+1}r_{j,t+1}$, the corrected sum of squares ($S_{r_i r_j}$) 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. 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 measure the negative of the absolute value of the difference in size, BE/ME and momentum percentile ranking across the two stocks in the pair ($SAME\_SIZE_{ij,t}$, $SAME\_BEME_{ij,t}$, and $SAME\_MOM_{ij,t}$ 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)
$SAME\_SIZE^*_{ij,t}$	-0.00028	-0.00007	0.00925	0.01143
	(-4.77)	(-1.64)	(6.72)	(7.36)
$SAME\_BEME^*_{ij,t}$	0.00009	0.00001	0.00264	0.00319
	(2.30)	(0.85)	(5.53)	(5.75)
$SAME\_MOM^*_{ij,t}$	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 1.7: Connected Comovement: Additional Controls and Decomposition

This table reports Fama-McBeth estimates of monthly cross-sectional regressions forecasting the realized cross-product of returns, $r_{i,t+1}r_{j,t+1}$, the daily return Fisher correlation (ρ_{Fisher}), and the cross products of the return components (cash-flow-news and discount-rate-news), $N_{CF,i,t+1}N_{CF,j,t+1}$, $-N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$, and $N_{DR,i,t+1}N_{DR,j,t+1}$ for the sample of stocks defined in Table 1.1. We estimate

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

where $\mathbf{y}= [r_{i,t+1}r_{j,t+1}; \quad \rho_{Fisher}; \quad N_{CF,i,t+1}N_{CF,j,t+1}; \quad -N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}; \quad N_{DR,i,t+1}N_{DR,j,t+1}]$. The return news components are extracted using the return VAR estimates shown in Table 1.2 and the methodology documented in the Appendix. We estimate the same equation as in Table 1.5, 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 both belong to the S&P 500 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 Table 1.2 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)
$SAME\_SIZE^*_{ij,t}$	-0.00023	0.01430	-0.00013	-0.00007	-0.00001
	(-4.00)	(9.10)	(-3.75)	(-1.30)	(-0.88)
$SAME\_BEME^*_{ij,t}$	0.00006	0.00189	0.00007	-0.00004	0.00002
	(1.94)	(4.13)	(3.75)	(-2.94)	(5.27)
$SAME\_MOM^*_{ij,t}$	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)
$D_{INDEX ij,t}$	-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 1.8: Alternative Measures of Connectedness

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 1.1. 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 $-N_{DR,i,t+1}N_{CF,j,t+1}-N_{DR,j,t+1}N_{CF,i,t+1}$ realized in month 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 1.5) and our specification that includes the Chen, Chen, and Li variables (Panel B, as in Table 1.7). As measures of common ownership, we use the number of common owners, $F_{ij,t}$; the Total Net Assets of all common owners across the two stocks, $F_{ij,t}^{TNA}$; the total ownership by all common funds in dollars of the two stocks scaled by the total market capitalization of the two stocks, $F_{ij,t}^{\%CAP}$; and the total ownership by all common funds in dollars of the two stocks scaled by the Total Net Assets of all common owners, $F_{ij,t}^{\%TNA}$. 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.

	Panel A: Benchmark			Panel B: 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%
$F_{ij,t}^{TNA*}$	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%
$F_{ij,t}^{\%CAP*}$	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%
$F_{ij,t}^{\%TNA*}$	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%

Table 1.9: Connected Comovement: 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 1.1. We estimate

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

where $\mathbf{y} = [r_{i,t+1}r_{j,t+1}]$. Panel A only considers a subset of these variables that are used in the regression in Table 1.5. Panel B estimates the full regressions specification. 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. In each Panel, we enhance the particular specification by interacting the common fund variable with dummies for the ranking of the pair based on quarterly independent sorts (as of time t) 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: $r_{i,t+1}r_{j,t+1}$

Benchmark controls of Table 1.5 included but not shown

$b_{f-k,l}$ estimates		Size of the pair (k)					
		Low	2	3	4	High	Low - High
	Low	0.00081	0.00075	0.00065	0.00051	0.00046	0.00034
		(4.50)	(5.30)	(5.85)	(5.67)	(5.05)	(2.54)
Total	2	0.00063	0.00059	0.00054	0.00043	0.00041	0.00022
net		(4.99)	(5.45)	(5.43)	(4.98)	(4.78)	(2.96)
flow	3	0.00068	0.00066	0.00061	0.00049	0.00045	0.00023
from		(4.28)	(4.60)	(4.70)	(4.50)	(4.33)	(2.51)
common	4	0.00065	0.00058	0.00055	0.00042	0.00035	0.00029
funds		(5.91)	(6.20)	(6.93)	(6.20)	(5.17)	(4.37)
	High	0.00119	0.00097	0.00074	0.00060	0.00048	0.00071
		(5.99)	(5.74)	(6.57)	(6.18)	(5.59)	(4.71)
	Low - 3	0.00013	0.00009	0.00004	0.00002	0.00002	
		(0.78)	(0.77)	(0.40)	(0.29)	(0.39)	
	High - 3	0.00051	0.00030	0.00014	0.00010	0.00003	
		(3.53)	(2.27)	(1.66)	(1.38)	(0.57)	

PANEL B: Dependent var: $r_{i,t+1}r_{j,t+1}$

All controls of Table 1.7 included but not shown

$b_{f-k,l}$ estimates		Size of the pair (k)					
		Low	2	3	4	High	Low - High
	Low	0.00077	0.00077	0.00069	0.00057	0.00050	0.00027
		(4.70)	(5.29)	(5.84)	(5.81)	(5.47)	(1.93)
Total	2	0.00060	0.00059	0.00057	0.00049	0.00046	0.00014
net		(5.47)	(6.27)	(6.15)	(5.93)	(5.54)	(1.64)
flow	3	0.00064	0.00062	0.00059	0.00052	0.00049	0.00016
from		(5.25)	(5.94)	(5.97)	(5.51)	(5.00)	(1.90)
common	4	0.00064	0.00058	0.00059	0.00050	0.00043	0.00021
funds		(7.07)	(8.17)	(8.49)	(7.55)	(6.13)	(3.37)
	High	0.00120	0.00100	0.00081	0.00070	0.00057	0.00063
		(5.95)	(6.10)	(7.02)	(6.39)	(5.53)	(4.42)
	Low - 3	0.00013	0.00015	0.00010	0.00005	0.00002	
		(0.92)	(1.36)	(1.42)	(1.07)	(0.53)	
	High - 3	0.00056	0.00038	0.00022	0.00018	0.00008	
		(3.88)	(3.12)	(3.38)	(2.74)	(1.97)	

Table 1.10: 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 $(r_{p,t} - r_{f,t})$, the equal-weight excess return on the portfolio of the stocks associated with the particular trading strategy, on four and five factors. Panel A reports α_4, four factor alphas (Carhart alphas) and Panel B reports α_5 five factor alphas (Carhart plus short-term reversal). In each panel, we also report the average returns on 1) a connected strategy, $CS1$, which buys the low own return / low connected return portfolio and sells the high own return / high connected return portfolio and 2) a second connected strategy, $CS2$, which buys the average (across the own return quintiles) low connected return portfolios and sells the average (across the own return quintiles) high connected return portfolios.

PANEL A: FOUR FACTOR ALPHAS								
		Connected portfolio						
		Low	2	3	4	High	L - H	Avg L-H
	Low	0.0042	0.0046	0.0044	0.0036	0.0008	0.0034	
		(2.97)	(3.97)	(3.82)	(2.93)	(0.59)	(1.81)	
	2	0.0053	0.0040	0.0029	0.0029	0.0009	0.0044	
Own		(4.54)	(4.46)	(3.33)	(3.16)	(0.88)	(3.13)	
Return	3	0.0037	0.0024	0.0015	0.0005	0.0000	0.0036	0.0036
		(3.41)	(2.72)	(1.84)	(0.54)	(0.01)	(2.81)	(3.01)
	4	0.0028	0.0000	-0.0006	-0.0010	-0.0014	0.0042	
		(2.35)	(0.05)	(-.79)	(-1.3)	(-1.7)	(3.05)	
	High	0.0004	-0.0004	-0.0024	-0.0027	-0.0022	0.0026	
		(0.32)	(-.40)	(-2.8)	(-2.9)	(-2.0)	(1.66)	
	L - H	0.0038	0.0050	0.0068	0.0063	0.0030	0.0064	
		(2.43)	(3.73)	(4.89)	(4.55)	(1.95)	(3.37)	

PANEL B: FIVE FACTOR ALPHAS								
		Connected portfolio						
		Low	2	3	4	High	L - H	Avg L-H
	Low	0.0040	0.0043	0.0041	0.0031	0.0005	0.0035	
		(2.84)	(3.73)	(3.54)	(2.53)	(0.39)	(1.85)	
	2	0.0053	0.0038	0.0027	0.0027	0.0007	0.0046	
Own		(4.50)	(4.29)	(3.10)	(2.91)	(0.65)	(3.29)	
Return	3	0.0035	0.0022	0.0013	0.0003	-0.0002	0.0037	0.0037
		(3.25)	(2.42)	(1.60)	(0.33)	(-.22)	(2.83)	(3.04)
	4	0.0027	-0.0003	-0.0009	-0.0011	-0.0017	0.0044	
		(2.25)	(-.27)	(-1.1)	(-1.4)	(-2.1)	(3.18)	
	High	0.0000	-0.0009	-0.0026	-0.0028	-0.0023	0.0023	
		(0.01)	(-.88)	(-3.0)	(-2.9)	(-2.1)	(1.46)	
	L - H	0.0040	0.0052	0.0067	0.0059	0.0028	0.0063	
		(2.56)	(3.87)	(4.78)	(4.21)	(1.82)	(3.30)	

Table 1.11: The Connected Strategy and Liquidity Risk

This table measures the loadings of the connected stock trading strategy on two common liquidity factors as well as on time effects. We study the connected strategy, $CS1$, formed in Table 1.10, which buys the low own return / low connected return portfolio and sells the high own return / high connected return portfolio so that its return is $r_{CS1}= r_{low\ own\ /\ low\ connected} - r_{high\ own\ /\ high\ connected}$. We regress r_{CS1} 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. Column 1 report loadings of our connected strategy on the five factors used in Table 10. Columns 2 and 3 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 4 to 6 include the PS liquidity factor, a trend, and quarterly seasonal dummies as additional explanatory variables, over the period June 1980 to December 2008.

	Dependent Variable: Connected Strategy					
	1	2	3	4	5	6
Alpha	0.0063	0.0063	0.0062	0.0063	0.0107	0.0109
	(3.30)	(3.28)	(2.87)	(3.28)	(2.95)	(3.02)
PS_INNOV		0.0638		0.0630		0.0708
		(2.03)		(2.00)		(2.23)
SADKA_PV			0.3564			
			(0.95)			
RMRF	-0.0081	-0.0377	0.0350	-0.0392	-0.0048	-0.0372
	(-0.16)	(-0.75)	(0.63)	(-0.78)	(-0.10)	(-0.74)
SMB	-0.3664	-0.3711	-0.4150	-0.3707	-0.3501	-0.3549
	(-5.97)	(-6.07)	(-6.22)	(-6.05)	(-5.61)	(-5.71)
HML	-0.1797	-0.1907	-0.1208	-0.1920	-0.1621	-0.1746
	(-2.53)	(-2.69)	(-1.50)	(-2.70)	(-2.24)	(-2.42)
UMD	-1.0164	-1.0132	-1.0120	-1.0136	-1.0191	-1.0164
	(-22.32)	(-22.34)	(-20.29)	(-22.31)	(-22.34)	(-22.41)
ST_Reversal	0.0164	0.0218	0.0398	0.0215	0.0201	0.0255
	(0.28)	(0.37)	(0.62)	(0.37)	(0.34)	(0.44)
Trend				0.0000		
				(-0.44)		
Q1					-0.0065	-0.0064
					(-1.24)	(-1.22)
Q2					-0.0087	-0.0099
					(-1.70)	(-1.94)
Q3					-0.0029	-0.0027
					(-0.56)	(-0.53)
Obs	343	343	274	343	343	343
R^2	65%	66%	67%	66%	66%	66%

Table 1.12: Hedge Fund and Mutual Fund Exposure to the Strategy

This table measures the exposure of two CSFB hedge fund return indexes (all and long/short) as well as the value-weight average active mutual fund return (net of fees) to the connected strategy described in Table 1.10. We regress fund index returns in excess of the t-bill return on a constant, the connected strategy and either the eight Fung and Hsieh (2001, 2004) hedge fund factors or the Fama-French/Carhart model plus a short-term reversal factor. The time period is January 1994 to December 2008. Panel A's analysis uses as the additional explanatory variable the connected strategy ($CS1$) in Table 1.10 that buys the low own return and low connected return portfolio and sells the high own return and high connected return ($r_{CS1}= r_{low\ own\ /\ low\ connected}-r_{high\ own\ /\ high\ connected}$). Panel B's analysis uses as the additional explanatory variable the connected strategy ($CS2$) in Table 1.10 that buys the average (across the own-return quintiles) low connected return portfolio and sells the average (across the own-return quintiles) high connected return ($r_{CS2}=\overline{r_{low\ connected}}-\overline{r_{high\ connected}}$).

PANEL A: $CS1$

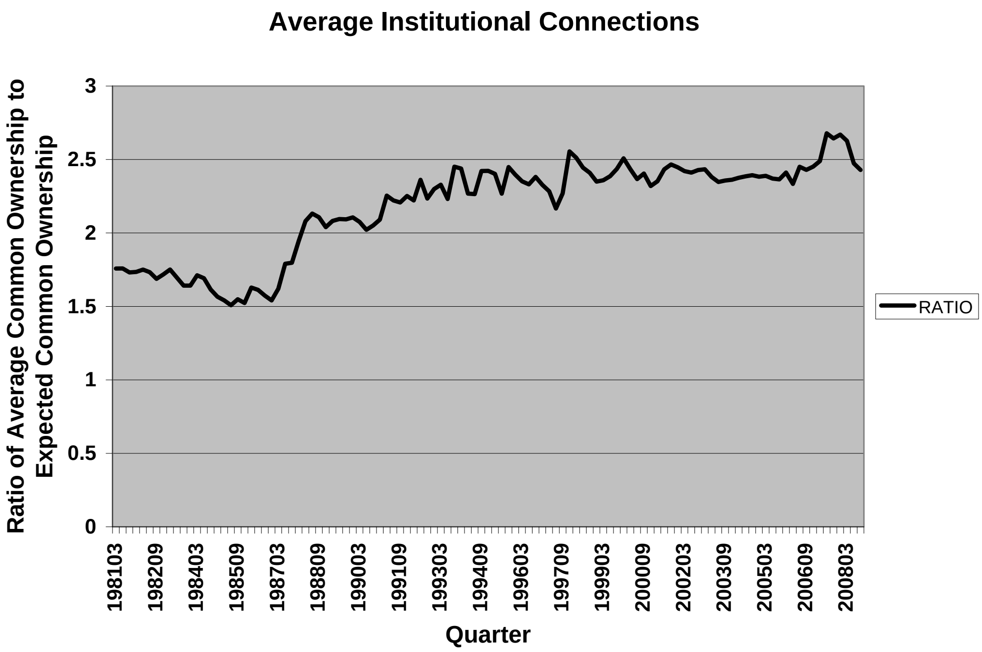
	HF ALL		HF LONG/SHORT		MF ALL (vw)
Alpha	0.0020	0.0022	0.0026	0.0012	-0.0013
	(1.70)	(2.06)	(2.73)	(1.09)	(-3.56)
r_{CS1}	-0.0658	-0.1114	-0.0817	-0.1707	-0.0265
	(-2.08)	(-6.09)	(-3.14)	(-9.44)	(-2.65)
RMRF	0.3794		0.5097		0.9934
	(13.07)		(21.34)		(108.3)
SMB	0.0852		0.1498		0.0562
	(2.31)		(4.95)		(4.84)
HML	0.0850		-0.0558		-0.0044
	(2.16)		(-1.72)		(-0.35)
UMD	0.0761		0.1223		-0.0071
	(1.84)		(3.60)		(-0.54)
ST Reversal	-0.0492		-0.0820		-0.0232
	(-1.67)		(-3.38)		(-2.50)
Bond-trend		-0.0226		-0.0084	
		(-2.96)		(-1.11)	
Currency-trend		0.0113		0.0050	
		(1.93)		(0.86)	
Commodity-trend		0.0131		0.0028	
		(1.63)		(0.35)	
Equity Market		0.1965		0.4140	
		(4.97)		(10.59)	
Size Spread		0.0629		0.2172	
		(1.88)		(6.56)	
Bond Market		-0.1235		-0.0090	
		(-3.41)		(-0.25)	
Credit Spread		-0.1816		0.0429	
		(-3.33)		(0.79)	
Emerging Market		0.0829		0.0897	
		(3.55)		(3.89)	
Obs	173	164	173	164	173
R^2	56%	60%	82%	76%	99%

PANEL B: $CS2$

	HF ALL		HF LONG/SHORT		MF ALL (vw)
Alpha	0.0019	0.0025	0.0024	0.0015	-0.0014
	(1.62)	(2.16)	(2.47)	(1.29)	(-3.89)
r_{CS2}	-0.1158	-0.1793	-0.0943	-0.2649	0.0006
	(-2.36)	(-4.44)	(-2.30)	(-6.24)	(0.03)
RMRF	0.3761		0.5071		0.9934
	(12.99)		(20.96)		(106.2)
SMB	0.0759		0.1564		0.0698
	(2.01)		(4.98)		(5.75)
HML	0.0815		-0.0523		0.0016
	(2.07)		(-1.59)		(0.12)
UMD	0.1015		0.1726		0.0209
	(3.31)		(6.74)		(2.11)
ST Reversal	-0.0528		-0.0854		-0.0236
	(-1.80)		(-3.48)		(-2.49)
Bond-trend		-0.0244		-0.0114	
		(-3.07)		(-1.36)	
Currency-trend		0.0097		0.0027	
		(1.60)		(0.41)	
Commodity-trend		0.0149		0.0058	
		(1.78)		(0.66)	
Equity Market		0.1834		0.3911	
		(4.43)		(9.00)	
Size Spread		0.0723		0.2347	
		(2.04)		(6.32)	
Bond Market		-0.1078		0.0128	
		(-2.79)		(0.31)	
Credit Spread		-0.1442		0.0982	
		(-2.51)		(1.62)	
Emerging Market		0.0811		0.0880	
		(3.31)		(3.42)	
Obs	173	164	173	164	173
R^2	56%	57%	81%	71%	99%

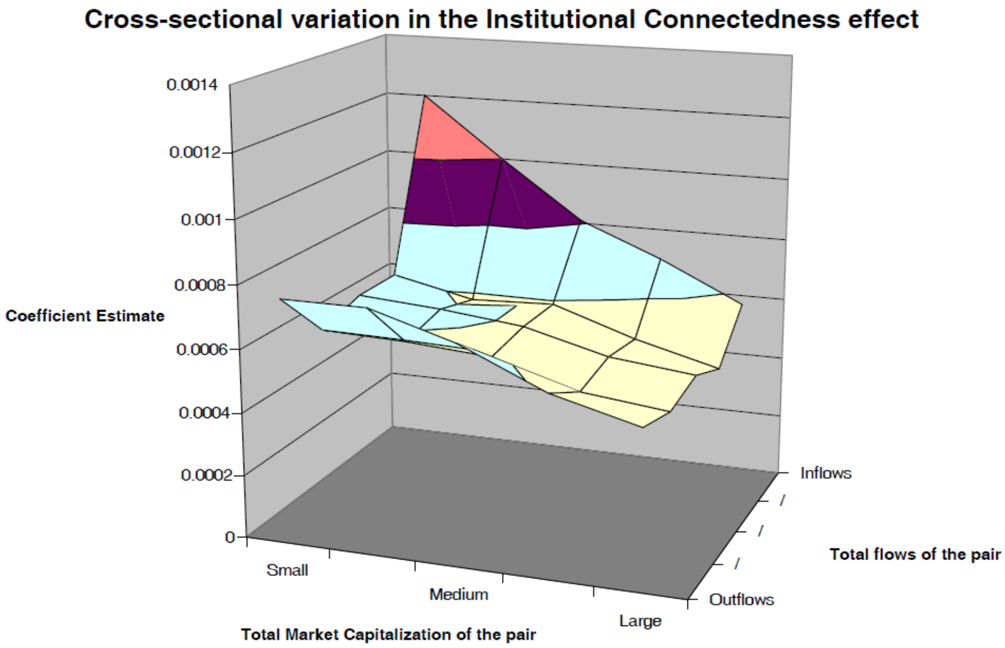
1.6 Figures

Figure 1.1: Average institutional connections



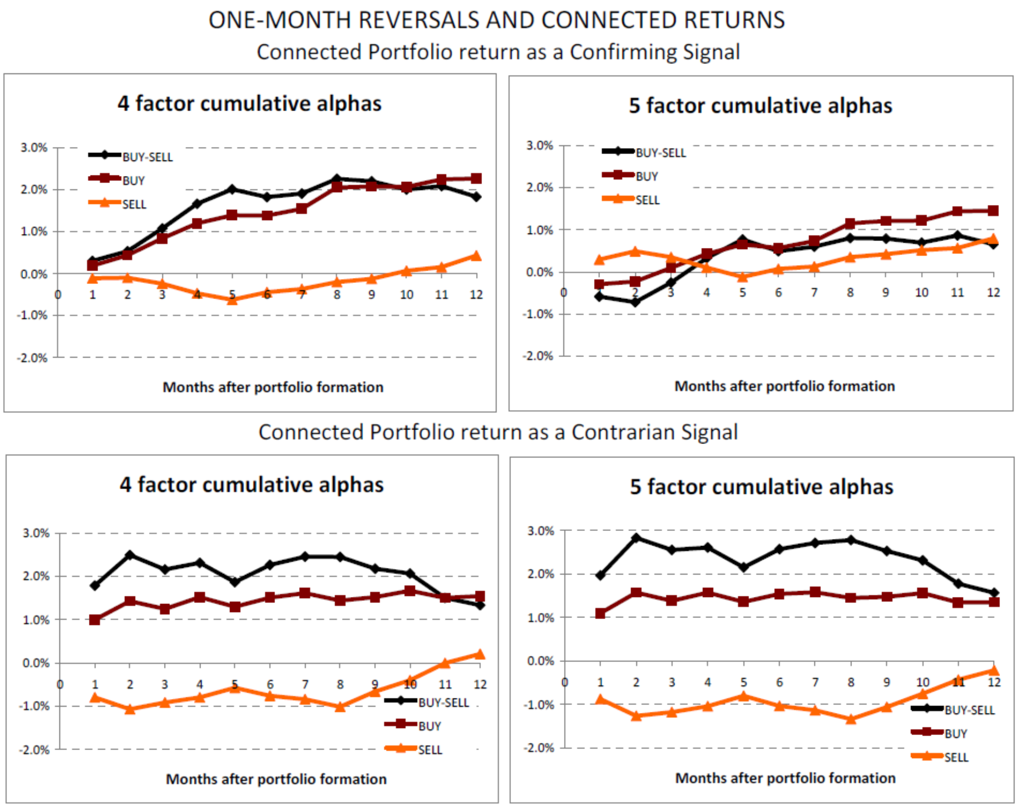
This figure plots the time-series evolution of the ratio of the average number of common funds per pair in each cross section of stock pairs to the average number of common funds per pair if all funds in that cross section held the same number of stocks as the average fund holds.

Figure 1.2: Cross-sectional variation in the institutional connectedness effect



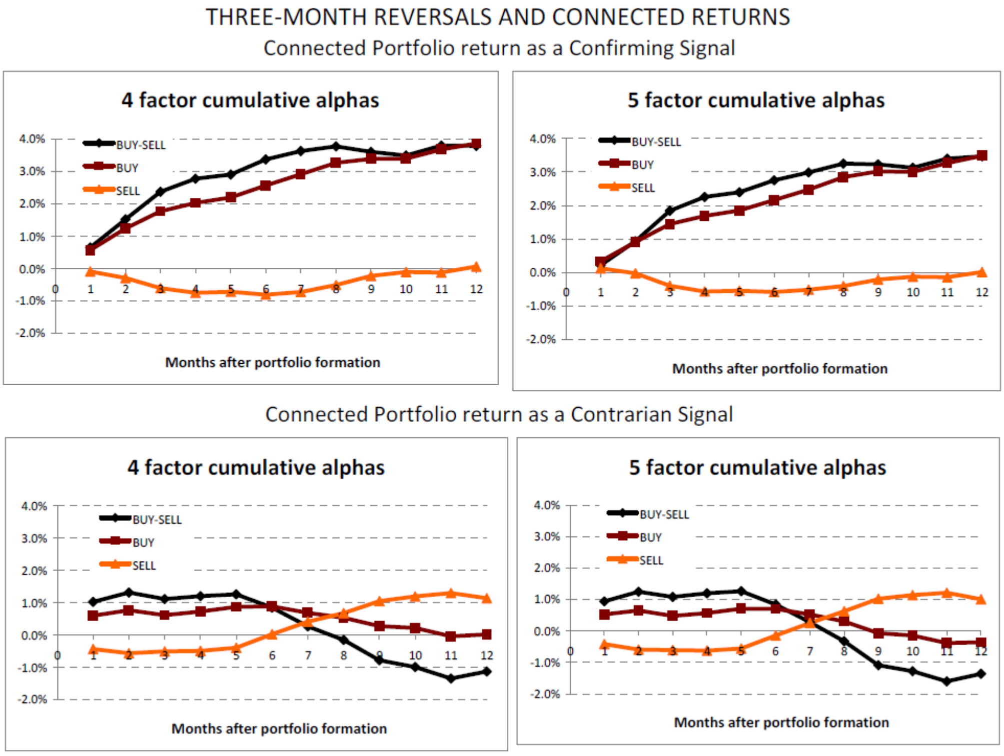
This figure plots the point estimates from Table 9 Panel A. In that table we interact the coefficient on the number of common funds per pair with dummies for the size of the pair of stocks and the total net flow into the common funds. Specifically, each quarter we sort pairs into quintiles based on their 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 1.3: One-month reversals and connected returns



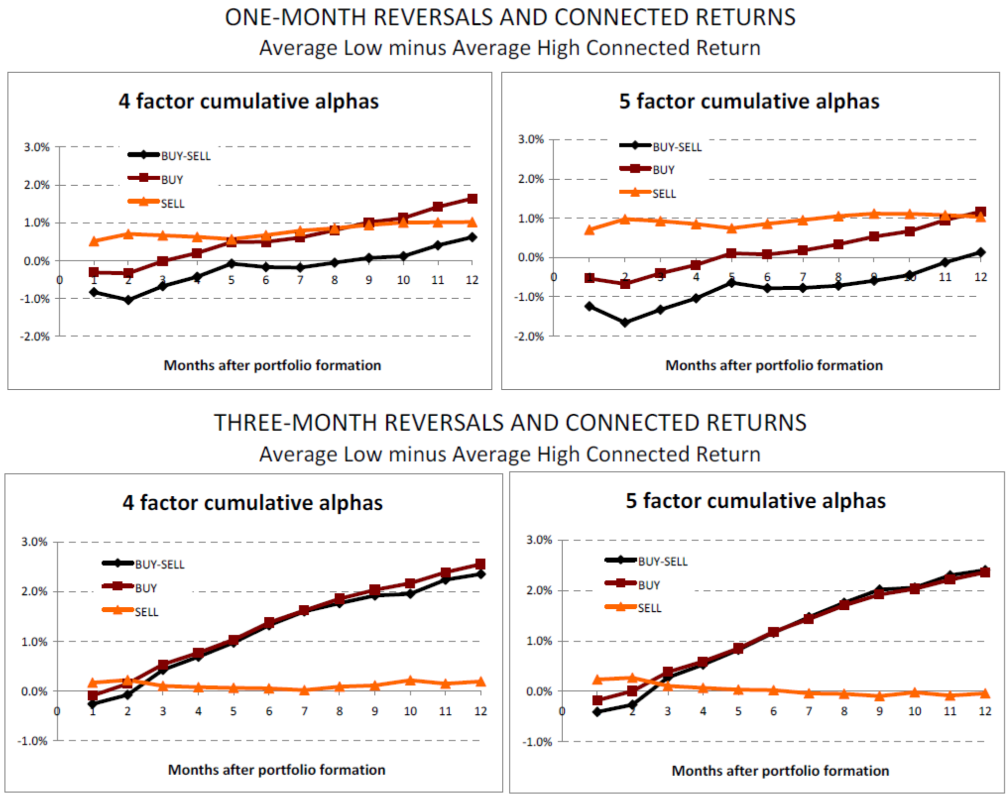
This figure graphs the abnormal performance of buy-and-hold strategies that trade the one-month reversal strategy conditional on the return on a stock's connected portfolio. Stocks are sorted into 25 portfolios based on independent quintile sorts on a stock's own one-month return and its one-month connected 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 Fama-French/Carhart four-factor model while the right side of the figure benchmarks returns against the Fama-French/Carhart model augmented with the one-month reversal factor.

Figure 1.4: Three-month reversals and connected returns



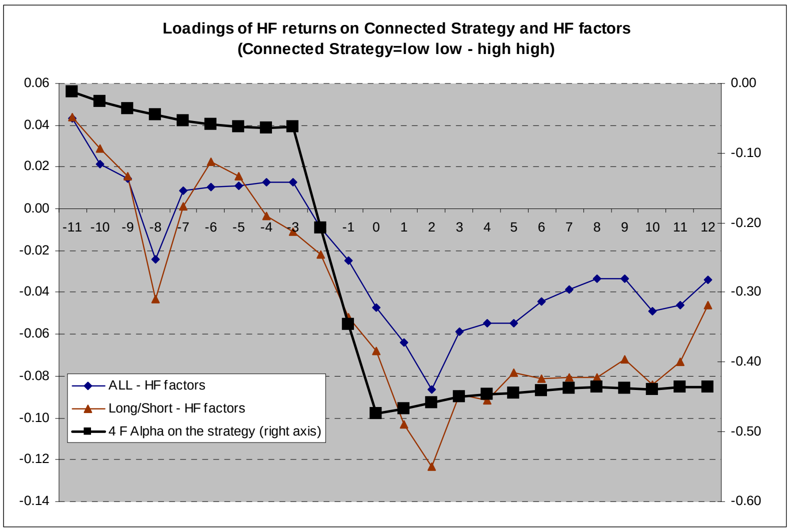
This figure graphs the abnormal performance of buy-and-hold strategies that trade a three-month reversal strategy conditional on the return on a stock's connected portfolio. Stocks are sorted into 25 portfolios based on independent quintile sorts on a stock's own three-month return and its three-month 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 Fama-French/Carhart four-factor model while the right side of the graphs benchmarks returns against the Fama-French/Carhart model augmented with the one-month reversal factor.

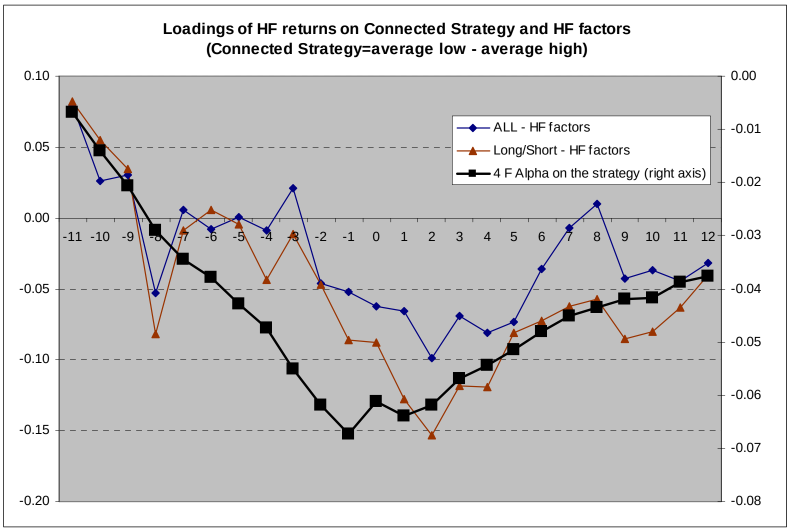
Figure 1.5: One-month reversals and connected returns, alternative strategy



This figure graphs the abnormal performance of buy-and-hold strategies that trade a one- and three-month reversal strategy based solely 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 and connected one-month (top two figures) or three-month (bottom two figures) returns. Each graph buys (sells) the average (across the own return quintiles) low (high) connected return portfolios. The left two graphs in the figure benchmark returns against the Fama-French/Carhart four-factor model while the right two graphs in the figure benchmark returns against the Fama-French/Carhart model augmented with the one-month reversal factor.

Figure 1.6: Loadings of hedge fund returns on connected strategies





This figure plots the loadings of hedge fund returns on the connected strategy in event time as well as the cumulative five-factor abnormal return. The top graph defines the connected strategy, $CS1$, which buys the low own return / low connected return portfolio and sells the high own return / high connected return portfolio so that its return is $r_{CS1} = r_{low\ own\ /\ low\ connected} - r_{high\ own\ /\ high\ connected}$. The second graph uses as the connected strategy, $CS2$, which buys the average (across the own return quintiles) low connected return portfolios and sells the average (across the own return quintiles) high connected return portfolios so that its return is $r_{CS2} = \overline{r_{low\ connected}} - \overline{r_{high\ connected}}$.

A Appendix

A.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:

$$\begin{aligned} r_{t+1} - \mathrm{E}_t\, r_{t+1} &= (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \quad (1.7) \\ &= N_{CF,t+1} - N_{DR,t+1}, \end{aligned}$$

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.

A.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 (1.7) 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 to make when 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

$$z_{t+1} = a + \Gamma z_t + u_{t+1}, \quad (1.8)$$

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 (1.8) generates the data, $t+1$ cash-flow and discount-rate news are linear functions of the $t+1$ shock vector:

$$\begin{aligned} N_{DR,t+1} &= e1'\lambda u_{t+1}, \\ N_{CF,t+1} &= (e1' + e1'\lambda)\, u_{t+1}. \end{aligned} \tag{1.9}$$

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.

A.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^e_M), 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, based on data available from Bob Shiller's website. 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 a version of the value spread introduced by Cohen, Polk, and Vuolteenaho (2003), but for small stocks (VS), which we construct using the data made available by Professor Kenneth French on his website. 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). As in Campbell and Vuolteenaho (2004), we generate intermediate values of VS by accumulating total returns on the portfolios in question.

Table 1.2 Panel A 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 1.2 Panel A 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, 1988a, 1988b, 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 1.2 Panel A 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.

A.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.[11] 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 $BM \equiv \log[(.9BE + .1ME)/ME]$.

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

[11] Following Fama and French (1993), 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 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 1.2 Panel B 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.

2 Cash-Flow Driven Covariation

2.1 Introduction

In standard finance models fundamentals drive asset prices. There is however a large body of the literature documenting departures of prices from fundamentals[12]. It is difficult to explain under the traditional paradigm market anomalies (e.g. momentum, reversal, value effect). Some of the evidence interpreted as favouring non-fundamental-based theories concerns index effects, both in first and second moments. For instance, Vijh (1994) and Barberis, Shleifer and Wurgler (2005) find that index additions are followed by an increase in covariation, and argue that this effect is not driven by fundamentals.

Index additions have been widely used as a quasi-natural experiment to distinguish between competing theories. For example, a number of papers show that there is a significant jump in price levels following index additions and deletions[13]. Much of these findings have been interpreted as evidence of non-fundamental-based theories. Some research, however, have challenged the interpretation of this effect. Dennis et al. (2003) for example argue that index additions are not fully information-free events, as they are followed by increases in earnings. While the interpretation of these effects in the first moments has been subject to debate among academics, changes in second moments (covariances) around index inclusions are widely accepted as evidence of non-fundamental-based theories[14].

In this paper I show that S&P 500 index inclusions are followed by changes in cash-flow covariances. I specifically take on the task of disentangling how much of the change in beta after an index addition corresponds to a fundamental effect and how much to a non-fundamental effect. I provide evidence of changes in cash-flow news' covariances after index additions using a two beta decompo-

[12] For instance, two recent papers survey the importance and implications of the limits of arbitrage for asset prices (Gromb and Vayanos, 2010, and Schwert, 2003).

[13] Starting with Harris and Gurel (1986), and Shleifer (1986), there are many studies that report significant changes in price levels. See Gromb and Vayanos (2010) for a survey on these effects.

[14] Barberis, Shleifer, and Wurgler (2005) say regarding Denis et al.: "Denis et al. (2003) find that index additions coincide with increases in earnings. [...] Perhaps more importantly, even if inclusions signal something about the level of future cash flows, there is no evidence that they signal anything about cash flow covariances".

sition. Following Campbell and Mei (1993), I decompose beta into discount-rate and cash-flow shocks of the individual firm with the market. I find that I cannot reject the hypothesis that all of the well-known change in beta comes from the cash-flow news component of a firm's return. As investors cannot directly influence fundamentals, these results challenge previous findings, as they are consistent with the change in beta being due to a selection effect.

The non-fundamental interpretation of the documented change in beta after an index inclusion is based on the key assumption that there is no change in fundamentals after index inclusions, nor a change in cash-flow covariances. S&P 500 index inclusions are considered as information-free events, because Standard and Poors clearly states that by choosing a firm to be added to the index they do not signal anything about the future fundamentals of that company. Consequently, a change in beta of stocks after the addition must reflect a change in discount-rates covariances, providing in this way evidence of friction- or sentiment-based comovement. My approach allows me to test whether the assumption actually holds.

Using vector-autoregressions (VARs), I break the returns of stocks added to the S&P 500 index into cash-flow and discount-rate components. That allows me to decompose the betas in two, one related to cash-flows and the other related to discount-rates of the event stocks. I find that, on average, the beta of the discount rate component does not change after an index inclusion, and that the beta of the cash-flow component does, and moreover accounts for the overall change in beta. I use a sample of index additions from September 1976 to December 2008.

I then study accounting-based fundamentals of included firms directly to reduce any concerns that the VAR-based results are sensitive to my particular specification. Using the return on equity as a direct measure of cash flows, this analysis confirms that post inclusion, the profitability of a company added to the index varies significantly more with the profitability of the S&P 500, and significantly less with the profitability of all non-S&P 500 stocks.

These results strongly suggest that Standard and Poors choices do not *trigger* or *cause* a change in betas after index inclusions, but rather it *selects* stocks that exhibit a growth in betas. S&P 500 Index is meant to be representative of the economy. Stocks are normally added following a deletion - which usually occurs due to mergers. The results are consistent with a story where Standard and Poors chooses stocks that are going to be more central to the economy, that will reflect the *state* of the economy, and thus that will have fundamentals more correlated to fundamentals of other representative firms in the economy. These results (where monthly frequency is used) complement the results found in Barberis et al. (2005). At higher frequencies, such as daily, the change in beta observed after an index addition reflects the change in speed at which information is incorporated into stocks. Due to market frictions, information is updated in S&P 500 stocks quicker than in non-S&P 500 stocks. In other words, the systematic risk does not change, what changes is how fast market news are embedded into stock prices. The results of the current paper, all computed at the monthly frequency (because a return decomposition is not feasible at higher frequencies), show that at lower frequencies there is indeed a change in the systematic risk of the stocks added to the index, and that this change is not causal, but reflects the evolutions of the fundamentals of event companies.

To better understand how the selection mechanism works, I develop a matching procedure, and measure the change in betas for companies that could have been added but were not. I find that matched stocks exhibit similar patterns in betas, and in some cases the difference in differences in betas is significant, as in previous literature. Using the beta decomposition, I find that the difference in differences is driven by cash-flow covariances, thus providing evidence of Standard and Poors signaling something about future cash-flow covariances. This finding is consistent with Standard and Poors' Committee being a better predictor of future cash-flow covariances and relevance in the economy than the basic and always imperfect matching algorithm that we employ.

Finally I explore the effect in different subsamples to uncover effects that might be hidden in the overall sample. First, subsampling in the time dimension, I find that the effect is stronger in the last part of the sample, and that the effect is driven by cash-flow covariances. Secondly, I study whether stocks with different characteristics differ in the change in beta experienced after inclusion. I divide the included firms into growth and value stocks, by comparing the cross-sectionally adjusted book-to-market ratios. Growth firms tend to be more intangible and more opaque, while value firms are more stable, if they are financially sound. Because the change in beta also reflects the size of the companies added, growth stocks should exhibit a higher change in beta than value stocks. Consistent with my prior, I find that the change in beta is higher for growth firms.

The results are robust to two other specifications of the VAR. Allowing for a more flexible and richer specification, I first estimate a second-order VAR, and show that the results are very robust to this new VAR. I also test a second alternative specification of the VAR, where firm-level and aggregate variables are state variables all together in a unique VAR, as opposed to the benchmark specification, where I estimate two different VARs, one for firm-level adjusted returns, and another one for market returns. Results are also very robust to the use of this alternative specification. The results are however ambiguous when I use the alternative cash-flow risk measure suggested by Da and Warachka (2009), based on an analyst earnings beta. In their paper they also show that the two ways of decomposing results (earnings beta and VAR) lead to different results.

This paper relates to two strands of the literature. On the one hand, it is related to the stock return comovement literature. It is well known that certain groups of stocks tend to have common variation in prices. These studies are divided in two groups: one supporting a fundamental view of comovement and the other supporting a friction- or sentiment-based view of comovement. The fundamentals-based view of comovement argues that stocks in certain groups (value or growth stocks) have common variation because of the characteristics of their cash-flows. For example, Fama and French (1996) argue that value stocks tend to comove because they are companies in financial distress and vulnerable to bankruptcy. Cohen, Polk, and Vuolteenaho (2009) find that the profitability of value stocks covaries more with market-wide profitability than that of growth stocks. The alternative view of comovement is the friction- or sentiment-based view, and argues that the stock market prices different groups of stocks differently at different times. For example, Barberis and Shleifer (2003) and Barberis, Shleifer and Wurgler (2005) argue that it is changes in investor sentiment that creates correlated movement in prices, although they lack common fundamentals. In this paper, I support the fundamentals-based view of comovement.

On the other hand, this paper is also related to the stream of the literature that studies the effects of index inclusions. A large body of literature explores the price effects of index inclusions. Some studies assume that S&P 500 inclusions are information-free events. Shleifer (1986) and Harris and Gurel (1986) find that there is an increase in price after an addition, but the effect dissipates after two weeks. They argue these findings are consistent with a perfectly elastic demand for stocks. Some authors claim that the index effect has a long-term impact on price. Wurgler and Zhuravskaya (2002) do not find a full reversal in prices, which suggests that the long-term demand curve is donward sloping. Other studies claim that S&P 500 inclusions are not information-free events. Dennis et al. (2003) find that a better monitoring improves the efficiency of managers of added companies, resulting in higher earnings after inclusions. Dhillon and Johnson (1991) find that the corporate bonds of companies added also respond to the listing announcement, and thus conclude that the announcement conveys new information about fundamentals. In this paper, I find supporting evidence of S&P 500 inclusions not being fully information-free events.

The remainder of the paper is organized as follows. In Section 2 I describe the decomposition of returns and betas. Section 3 shows the VAR framework and VAR estimations. In Section 4 I show the empirical results, and the robustness checks. Section 5 concludes.

2.2 Decomposing Stock Returns and Betas

The main purpose of this paper is to understand the sources of change in betas around S&P 500 inclusions, and the novelty of this paper is precisely to break return betas into discount-rate and cash-flow betas in the context of S&P 500 additions to distinguish between fundamentals and sentiment theories. In this Section I describe carefully how we can break betas into discount rate and cash-flow betas. Drawing from previous literature, I will first explain how returns are decomposed, and then I turn to apply this decomposition to betas.

2.2.1 Decomposing Returns

Following the Gordon growth formula, the price of a financial asset is expressed as the sum of its expected future cash flows, discounted to the present with a set of discount rates. The source of change in the price of the asset comes from either a change in the expected stream of cash flows, or from a change in the expected discount rates.

Decomposing returns in the context of index additions is useful because it allows me to distinguish between fundamentals and sentiment stories for two reasons. The first one is that investors cannot directly affect the fundamentals of a firm. As a consequence, any impact of investor sentiment in prices is made through the channel of discount rates. Changes in investor sentiment, thus, means that investors change the discount rates they apply to otherwise unchanged set of cash-flows. Secondly, the origin of a change in price matters for long-term investors, such as pension funds. If returns drop caused by an increase in discount rates, these investors are not too concerned, because this is partially compensated by better future investment opportunities. However, if the drop in current returns reflect a fall in the expected cash-flows, this loss is not compensated. A good example of this effect is the recent study by Campbell, Giglio, and Polk (2010), where they show how similar drops in aggregate returns can affect long-term investors very differently depending on the sources of these downturns.

To decompose returns, I follow the framework set up by Campbell and Shiller (1988a, 1988b). They loglinearize the log-return:

$$r_{t+1} = \log(P_{t+1} + D_{t+1}) - \log(P_t) \tag{2.1}$$

where r denotes log-return, P the price, and D the dividend. They approximate this expression with a first order Taylor expansion around the mean log dividend-price ratio, $(\overline{d_t - p_t})$, where lowercase letter denote log transforms. This approximation yields

$$\begin{aligned} r_{t+1} &\approx k + \rho p_{t+1} + (1-\rho) d_{t+1} - p_t \qquad (2.2) \\ \text{where} \quad \rho &\equiv 1/(1+\exp(\overline{d_t - p_t})) \\ k &\equiv -\log(\rho) - (1-\rho)\log(1/\rho - 1) \end{aligned}$$

In this approximation, the log sum of price and dividend is replaced by a weighted average of log price and log dividend.

We now solve iteratively equation 2.2, by taking expectations and assuming that $\lim_{j\to\infty} \rho^j (d_{t+j} - p_{t+j}) = 0$, and get

$$p_t - d_t = \frac{k}{1-\rho} + E_t \sum_{k=1}^{\infty} \rho^j [\Delta d_{t+1+j} - r_{t+1+j}] \tag{2.3}$$

This accounting identity states that the price dividend ratio is high when the expected stream of future dividend growth (Δd) is high or when expected returns are low.

Drawing from this result, Campbell (1991) develops a return decomposition based on the loglinearization. The results obtained in equation 2.3 are plugged into equation 2.2. Then, substracting the expectation of log return, we get

$$\begin{aligned} r_{t+1} - \mathrm{E}_t\, r_{t+1} &= (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=0}^{\infty} \rho^j \Delta d_{t+1+j} - (\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j} \\ &= N_{CF,t+1} - N_{DR,t+1}, \end{aligned} \tag{2.4}$$

where N_{CF} and N_{DR} denote news about future cash flows (future dividends), and news about future discount rates (i.e., expected returns) respectively. Unexpected stock returns are thus a combination of changes in expected future cash flows and expected future discount rates.

2.2.2 Decomposing Betas

If a stock's beta is defined as the correlation of the stock return with the market return, then we can break betas into different components using the return decomposition described above. Previous research has used the return decomposition shown in equation 2.4 to break systematic risk in different ways. Campbell and Mei (1993) decompose the returns on stock portfolios (sorted on size or industry) and compute the cash-flow and discount-rate news of each portfolio. They define two beta components, one measuring the sensitivity of cash-flow news of the portfolio with the market and the other measuring the sensitivity of discount-rate news of the portfolio with the market. The two beta components are the following:

$$\beta_{CFi,M} \equiv \frac{Cov_t(N_{i,CF,t+1}, r_{M,t+1})}{Var_t(r_{M,t+1})} \tag{2.5}$$

and

$$\beta_{DRi,M} \equiv \frac{Cov_t(N_{i,DR,t+1}, r_{M,t+1})}{Var_t(r_{M,t+1})} \tag{2.6}$$

These two beta components add up to the traditional market beta of the CAPM:

$$\beta_{i,M} = \beta_{CFi,M} + \beta_{DRi,M} \tag{2.7}$$

Unlike Campbell and Mei (1993), I will break the betas on individual stocks (those added to the S&P 500 index), rather than on stock portfolios.

2.3 A VAR framework

2.3.1 Measuring the components of returns

I use vector autoregressions (VARs) to measure the shocks to cash flows and to discount rates, following Campbell (1991) approach. The 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 2.4 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, I assume that the data are generated by a first-order VAR model

$$z_{t+1} = a + \Gamma z_t + u_{t+1}, \tag{2.8}$$

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.

Assuming that the process in equation (2.8) generates the data, $t+1$ cash-flow and discount-rate news are linear functions of the $t+1$ shock vector:

$$\begin{aligned} N_{DR,t+1} &= e1'\lambda u_{t+1}, \\ N_{CF,t+1} &= \left(e1' + e1'\lambda\right) u_{t+1}. \end{aligned} \tag{2.9}$$

where $e1$ is a vector with first element equal to unity and the remaining elements equal to zero. 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.

2.3.2 Aggregate VAR Specifications

For my analysis I need to break individual stock returns into cash-flow and discount-rate news. However, as pointed out by Vuolteenaho (2002), it is useful and accurate to carry out the decomposition in two steps. Because aggregate returns behave differently than firm-level returns, it is reasonable to estimate a VAR for market returns, using aggregate variables, and a VAR for firm-level market-adjusted returns, using firm-level variables. Consistent with Vuolteenaho (2002), I show in the last section that estimating a unique VAR for firm-level stock returns delivers similar results.

I first estimate an aggregate VAR, to predict market returns. In specifying the aggregate VAR, I include four variables, following Campbell and Vuolteenaho (2004). The data are all monthly, from December 1928 to May 2009.

The first element the VAR is the excess return on the market (r_m^e), calculated as the difference between the monthly log return on the CRSP value-weighted stock index (r_m) and the monthly log risk-free rate (r_f). I take the excess return series from Kenneth French's website[15]. The second element in the VAR 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[16]. 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

[15] http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

[16] This last variable is only available until 2002, from that year until the end of the series I compute the TY series as the difference between the yield on the 10-Year US Constant Maturity Bond (IGUSA10D) and the yield on the 1-Year US Constant Maturity Bond (IGUSA1D).

in the index. I take the price-earnings ratio series from Robert Shiller's website[17]. As in Campbell and Vuolteenaho (2004), I carefully remove the interpolation inherent in Shiller's construction of the variable to ensure the variable does not suffer from look-ahead bias. The final variable is the small-stock value spread (VS), which I 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). I generate intermediate values of VS by accumulating total returns on the portfolios in question.

The motivation for the use of these variables is the following. Term yield spread tracks the business cycle, as pointed out by Fama and French (1989), and there are several reasons why we should expect aggregate returns to be correlated to the business cycle. Second, if price-earnings ratio is high and expected earnings growth is constant, then long-run expected returns must be low, so we expect a negative coefficient of this variable in the VAR. Finally, the small-stock value spread is included given the evidence in Brennan, Wang, and Xia (2001) and others that relatively high returns for small growth stocks predict low aggregate returns in the market.

Table 2.1 reports the VAR model parameters for the aggregate VAR, estimated using OLS. Every 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 in Table 2.1 shows that all four of my VAR state variables have some ability to predict monthly excess returns on the market excess returns. Monthly market returns display momentum; the coefficient on the lagged market excess 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 1.6. As expected, the smoothed price-earnings ratio negatively predicts market excess returns, with t-

[17] http://www.econ.yale.edu/~shiller/data.htm

statistics of 3.41, 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 market excess returns with t-statistics of 2.16, consistent with Brennan, Wang, and Xia (2001), Eleswarapu and Reinganum (2004), and Campbell and Vuolteenaho (2004). The estimated coefficients, both in terms of signs and t-statistics, are consistent with previous research.

The remaining rows in Table 2.1 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.

2.3.3 Firm-level VAR Specification

After the estimation of an aggregate VAR, I now turn to estimate a firm-level VAR for market-adjusted returns. I implement the main specification of my 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, I unlever the stock by 10 percent; that is, I 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. My second state variable is the momentum of the stock (MOM), which I measure following Carhart (1997) as the cumulative return over the months $t-11$ to $t-1$. My final firm-level state variable is the log book-to-market equity ratio (I 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.

I 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[18]. I update BE/ME over

[18]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

the subsequent eleven months by dividing by the cumulative gross return from the end of May to the month in question. I 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, I 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, I first shrink the BE/ME towards one by defining $BM \equiv \log[(.9BE + .1ME)/ME]$.

The firm-level VAR generates market-adjusted cash-flow and discount-rate news for each firm and month. I 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 $r_{i,t}$, I subtract the log value-weight CRSP index return, 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, I estimate the coefficients of the firm-level VAR using WLS. Specifically, I multiply each observation by the inverse of the number of cross-sectional observation that year, thus weighting each cross-section equally. This ensures that my estimates are not dominated by the large cross sections near the end of the sample period. I 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 my results in a measurable way.

Parameter estimates, presented in Table 2.2, 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.

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.

2.4 Empirical Results

2.4.1 Data

I use S&P 500 index inclusions between September, 1976 and December 31, 2008. There are 745 inclusion events in the sample period. Following prior studies, I exclude the events where the included firm is a spin-off or a restructured version of a firm already in the index, if the firm is engaged in a merger or takeover around the inclusion event, or if the event occurs so close to the end of the sample that the data required for estimating post-event betas are not available.

I do not consider deletion events in this study for two main reasons. Firs, most of the deletions from the S&P 500 (over 80%) are derived from a spin-off, mergers or restructuring. The second reason is that the evidence of beta shifts followed by deletions reported in the literature is smaller and less significant than that of additions.

I use monthly and quarterly data, from CRSP and Compustat. The analysis is done at the monthly frequency, because the return decomposition is done monthly. Higher frequency return decomposition is not considered, because the state variables used in the VAR are based on accounting variables, available at low frequencies.

Data for inclusion events comes from two sources: CRSP Index file, provided by Standard and Poors, and Jeffrey Wurgler's website. From 1976 to 2000 I use Jeffrey Wurgler's sample (590 additions), that includes information on whether the addition is related to mergers or spin offs. From 2001 to 2008 I obtain the data from CRSP Index file (155 additions), and manually investigate confounding events, using Nexis, Wall Street Journal, the companys' websites, Google.com, and Wikipedia. I exclude 33 additions that are related to mergers or spin-offs. I also require the additions to have enough data on the return decomposition.

2.4.2 Changes in Betas in a VAR Framework

2.4.2.1 Benchmark case I first conduct a basic bivariate regression where I measure the change in beta of the event stocks with respect to the S&P 500 return, controlling for the non S&P 500 return. I do this following the empirical approach of Barberis, Shleifer, and Wurgler (2005). They conjecture that controlling for the return of the "exiting" group (all non S&P 500 stocks) gives more power to distinguish between fundamentals and friction- or sentiment-based views.

I build a panel of all the event stocks, using a window of 36 months before and 36 months after the addition. I include the interaction of $r^e_{SP,t}$ and $r^e_{nSP,t}$ with a dummy variable I_{it} that takes value 1 if the stock is included in the index. The subscript t reflects event time (months around the inclusion), not calendar time. The equation I estimate is therefore the following:

$$r^e_{i,t} = \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \quad (2.10)$$

The coefficients of the interactions $I_{it} * r^e_{SP,t}$ and $I_{it} * r^e_{nSP,t}$ ($\Delta\beta_{SP}$ and $\Delta\beta_{nSP}$ respectively) reflect the average changes in betas after the addition to the S&P 500 index has taken place. The excess return on the S&P 500 index, r^e_{SP}, is computed as the difference between the monthly return on the S&P 500 index, obtained from the CRSP Index File, and the monthly riskfree rate, obtained from Professor Kenneth French's website. The return r^e_{nSP} are excess returns on a capitalization-weighted index of the non-S&P 500 stocks in the NYSE, AMEX, and Nasdaq, and are inferred from the following identity:

$$r_{M,t} = \left(\frac{CAP_{M,t-1} - CAP_{SP,t-1}}{CAP_{M,t-1}} \right) r_{nSP,t} + \left(\frac{CAP_{SP,t-1}}{CAP_{M,t-1}} \right) r_{SP,t} \quad (2.11)$$

where total capitalization of the S&P 500 (CAP_{SP}) is from the CRSP Index on the S&P 500 Universe file. Returns on the value-weighted CRSP NYSE, AMEX, and Nasdaq index (r_M) and total capitalization (CAP_M) are from the CRSP Stock Index file.

The constant in this regression has the i subscript, which means that I include firm dummies. It is reasonable to assume that the alphas for each event stock are different. Moreover, if two additions are close together in time, there can be overlap in the time periods covered by the regressions associated with each event. To account for this cross-sectional autocorrelation, I cluster standard errors by time (month).

Table 2.3 shows the results for this regression. Consistent with previous literature (Barberis, Shleifer, and Wurgler, 2005), I find that beta with respect to S&P 500 returns jumps and beta with respect to non S&P 500 returns falls, both significantly. The second row displays the average change in S&P 500 beta, $\Delta\beta_{SP}$, 0.425, accurately estimated with a t-stat of 6.25. The fourth row shows the average change in non S&P 500 beta, $\Delta\beta_{nSP}$, with the coefficient -0.291, estimated with a a t-stat of 4.59.

2.4.2.2 Cash-flow and discount-rate betas The results reported in Table 2.3, in line with those found by Barberis et. al, have been interpreted as evidence of friction- or sentiment-based comovement. The argument is the following. Standard and Poors state clearly that in choosing a company to be included in the index, they do not signal anything about the future performance of the company. As a consequence, any change in the betas of companies added to the index should be attributed to sentiment, because fundamentals have not changed.

Sentiment- or friction-based theories predict that the increase in beta is due to an induced common factor in the discount rates. Investors cannot affect directly the fundamentals (cash-flows) of a firm. However, they can apply similar discount rates to stocks in the same group, thus inducing an excess comovement.

Examining the components of the change in beta follows naturally from this argument. If the excess comovement is driven by sentiment- or friction-based reasons, then the observed change in beta should be coming from a change in discount rate betas, and we should not observe a change in cash flow covariances. If, however, the change is driven by cash-flow covariances, then this is support for a fundamentals-based view of comovement.

To implement this test, I simply substitute the excess returns of event stocks, $r^e_{i,t}$, for their cash-flow news ($N_{iCF,t}$) and (negative of) discount-rate news ($-N_{iDR,t}$) in the left-hand side of equation 2.10:

$$-N_{iDR,t} = \alpha_i + \beta_{SP}^{DRb} r_{SP,t}^e + \beta_{nSP}^{DRb} r_{nSP,t}^e + \Delta\beta_{SP}^{DR} I_{it} r_{SP,t}^e + \Delta\beta_{nSP}^{DR} I_{it} r_{nSP,t}^e + \varepsilon_{i,t} \tag{2.12}$$

and

$$N_{iCF,t} = \alpha_i + \beta_{SP}^{CFb} r_{SP,t}^e + \beta_{nSP}^{CFb} r_{nSP,t}^e + \Delta\beta_{SP}^{CF} I_{it} r_{SP,t}^e + \Delta\beta_{nSP}^{CF} I_{it} r_{nSP,t}^e + \varepsilon_{i,t} \tag{2.13}$$

so that I can identify the changes in beta due to discount rates, and those due to cash-flows. This decomposition implies that the overall change in beta with respect to S&P 500 (and similarly with non S&P 500 stocks), is approximately equal to the sum of changes in cash-flow betas and discount rate betas:

$$\begin{aligned} \Delta\beta_{SP} &\approx \Delta\beta_{SP}^{DR} + \Delta\beta_{SP}^{CF} \\ \Delta\beta_{nSP} &\approx \Delta\beta_{nSP}^{DR} + \Delta\beta_{nSP}^{CF} \end{aligned} \tag{2.14}$$

Table 2.4 shows the changes in cash-flow and discount rate betas. The first column replicates the benchmark column of table 2.3. The second and third columns show the results for the change in the different beta components. The change in discount rate beta with respect to the S&P 500 is an insignificant -0.008 (second row, second column), and 0.049 with respect to the non S&P 500 stocks, whereas the changes in cash-flow betas are 0.391 and -0.286 (for S&P 500 and non S&P 500 respectively), accurately estimated with t-stats of 6.15 and 4.62. This result strongly supports the idea that, at the monthly frequency, sentiment- or friction-based comovement is negligible if not inexistent.

Figure 2.1 shows the evolution of average betas around the inclusion event. Rolling regressions are estimated with windows of 36 months from month -36 to month $+72$. In the top panel we observe the evolution of the overall average betas. S&P 500 betas increase significantly after inclusion, and non S&P 500 decrease after inclusion. Below, in the central panel, rolling average discount rate betas are plotted, showing a very mild pattern of variation. Finally, in the bottom panel, we see how all the action in the change in beta is originated in the cash-flow betas.

2.4.3 Results from a direct approach

In this subsection I avoid the need for a VAR estimation, and thus show that my results do not depend on the VAR specification nor on the state variables used in the VAR. The main result arising from the previous section is that the changes in overall betas with S&P 500 and non S&P 500 returns come from cash-flow betas. In other words, I have found evidence that the fundamentals of stocks added to the S&P 500 index tend to comove more with fundamentals of the S&P500 after inclusion than before.

I use the return on equity (roe_{it}) to proxy for firm-level cash flow fundamentals, as done previously in the literature (Cohen, Polk, and Vuolteenaho, 2003, 2009). The specification is very simple: I regress the individual roe_{it} on the aggregate return on equity for the S&P 500 ($roe_{SP,t}$), on the aggregate return on equity for the rest of the market ($roe_{nSP,t}$), and on the interaction of these two variables with a dummy variable I_{it} that is equal to 1 if the stock is in the index and equal to 0 if it is not. The hypothesis is that if there is a change in the cash-flow covariances of the event stocks with the S&P 500 index, then I should observe a positive coefficient for the first interaction term ($I_{it}roe_{SP,t}$) and a negative coefficient for the second interaction term ($I_{it}roe_{nSP,t}$). The specification is then

$$roe_{i,t} = \alpha_i + \beta^b_{SP} roe_{SP,t} + \beta^b_{nSP} roe_{nSP,t} + \Delta\beta_{SP} I_{it} roe_{SP,t} + \Delta\beta_{nSP} I_{it} roe_{nSP,t} + \varepsilon_{i,t}$$

where $roe_{i,t}$ is the return on equity, defined as $roe_{i,t} = \log(1 + NI_t/BE_{t-1})$ where NI is net income and BE book equity, in t and $t-1$ respectively. To avoid extreme observations, $roe_{i,t}$ is winsorized between -1 and 2 (on a given quarter, the return on equity cannot be lower than -100% or higher than 200%). $roe_{SP,t}$ and $roe_{nSP,t}$ are calculated as the log of 1 plus the sum of NI_t over the sum of BE_{t-1}, for all December fiscal year end stocks in each group of S&P 500 and non S&P 500 stocks. As in the previous analyses, I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

I run a pooled-OLS quarterly regression. Results are presented in table 2.5. The results confirm my findings in the VAR approach. When a stock is not in the index, its beta with S&P 500 return on equity is 0.227 and its beta with the rest of the market return on equity is 0.716, with both coefficients estimated precisely with t-statistic above 3. However, once the stock has been added to the index, the betas turn to 0.488 and 0.211 for S&P 500 and rest of the market return on equities.

2.4.4 Matched stocks

The results from the VAR and from the direct approach strongly suggest that S&P 500 additions do not trigger a change in betas, rather, it *selects* stocks that exhibit a growth in betas. In other words, the observed change in beta of stocks added to the S&P 500 is not a *consequence* of being added, but rather, a *motive* for being added. S&P 500 index is meant to be representative of the economy, normally composed by large firms. The results are consistent with a story where Standard and Poors chooses stocks that are going to be more central to the economy, by having fundamentals more correlated with the fundamentals of other representative companies.

A natural exercise that helps to distinguish between causality and selection is a matching procedure. We can identify stocks of similar characteristics than those added to the S&P 500, but that happened not to be added. If S&P 500 additions are triggering or causing a change in beta, then event stocks should exhibit a change in betas coming from the discount rates, whereas matched stocks should not. If, however, it is Standard and Poors that is selecting stocks from certain sector and characteristics, then we would observe similar patterns of comovement in matched stocks as well.

Following Barberis et al., for each event stock I search for a matching stock similar in size and industry. I choose a stock in the same size decile at the moment of inclusion and 36 months before inclusion. I first match at the SIC4 level. If no match can be found, I allow the matched stock to be in the same SIC3 level. If no match is found, I then go back to SIC4 level and allow the matched stock to be within one size decile at inclusion, then within one size decile 36 months before inclusion. If no match can be found, I repeat the size allowance for SIC3 level, and then for the SIC2 level. I finally repeat the same algorithm for allowance of two size deciles at inclusion and then 36 months before inclusion.

Table 2.6 shows the results of the changes in beta using matched stocks. I find that matched stocks exhibit similar patterns in betas, as matched stocks also experience a significant change in beta with respect to S&P 500 returns, of 0.261. The crucial result in this table is that the difference in difference in betas, though mildly significant (0.165 with a t-stat of 1.91), it all comes from the cash-flow component: 0.158 with a t-stat of 2. This is both evidence of Standard and Poors signaling something about future cash-flow covariances, and of Standard and Poors' Committee being a better predictor of future cash-flow covariances and relevance in the economy than the basic and always imperfect matching algorithm that we employ.

Figure 2.2 shows the evolution of rolling average betas (for the overall betas, and their discount-rate and cash-flow components). The top panel shows the betas for the event firms (those included in the S&P 500), and the bottom panel shows the evolution of betas for matched firms (firms that could have been included in the index, but were not).

2.4.5 Reconciling with Barberis et al.

How do these results compare to those of Barberis et al.? They provide evidence of an excess-comovement coming from sentiment, and in this paper I provide evidence of a cash-flow driven comovement after index inclusions. In this subsection I explicitily compare both results to better understand how they relate to each other.

Barberis et al. provide empirical evidence supportive of three sentiment- or friction-based views of comovement. The category view, proposed by Barberis and Shleifer (2003), argues that investors, in order to simplify portfolio decisions, allocate funds at the category level, instead of asset level. Thus if there are noise traders with correlated sentiment, and they are effective in affecting prices, they create an excess comovement into each by moving funds from one to another category. Habitat view is based on the fact that many investors limit their investment universe to a preferred habitat, due to transaction costs, or lack of information. This in turns creates a common factor in the returns of these assets that is uncorrelated to fundamentals. The information difussion view stems from the fact that due to market frictions, the information is incorporated quicker into the prices of some stocks than others.

The two main contributions of their paper with respect to Vijh (1994) are as follows. They first extend the sample and show that the results are stronger in the recent period. Secondly they run bivariate regressions to enhance the power of the tests, by controlling in the regressions for non-S&P 500 returns. This methodology follows from the first two views of sentiment-based comovement: when a stock joins a group of stocks, the comovement of the stock with the new group should go up (as seen in Vijh), but also, and this is the novel approach, the comovement of the stock with the group to which it belonged (the *leaving* group), should drop.

They show that the evidence of excess-comovement after index inclusions is strong when using daily data, and becomes weaker when using lower frequencies of the data. Results for weekly and monthly data, although present, are less powerful than those using daily data. So the frequency used in the analysis matters. To understand how the three views contribute to the effect, Barberis et al. add a final section in the paper where they repeat the daily analysis using Dimson betas: using five leads and five lags of the right hand side variables, namely, S&P 500 index and non-S&P 500 index. They find that most of the effect dissappears when controlling for Dimson betas. Some of the effect remains in the univariate analysis, however statistical significance dissappears in the bivariate analysis, which is, in turn, the novel methodology they propose to enhance the power of the tests. Results are also shown only for event stocks, suggesting that difference in differences for matched stocks is not significant.

In this paper I show that there is a significant change in the covariances after index inclusions, and that such a change comes from the cash-flow component of the return covariance. I only use monthly frequency, as a return decomposition at higher frequencies is not feasible given the frequency of the variables that predict returns.

The results of Barberis et al., with especial emphasis on the Dimon betas analysis, together with my results strongly suggest that at high frequencies, the change in beta reflects the friction-based view of information difussion. Stocks in the S&P 500 index incorporate information quicker than stocks outside the S&P 500 index. In other words, an inclusion in the index changes the speed at which information is incorporated, but it does not change the systematic risk of the stocks added to the index. At lower frequencies, however, when we observe a change in the systematic risk of a stock added to the index, this change does not reflect a change in the speed of information incorporation (a causal effect triggered by the inclusion), but rather it reflects the evolution of the fundamentals of the stock added to the Index. This evolution in fundamentals is also present in matched stocks that were not added to the Index.

2.4.6 Robustness to different subsamples

2.4.6.1 Subsample in the time dimension I explore the effect in different time subsamples to uncover effects that might be hidden in the full-sample period. Previous research has found that the change in beta after index additions has grown over time. Consistent with those findings, I find that the effect is stronger in the last part of the sample. This analysis, shown in table 2.7, reflects three findings. Firstly, the effect of the change in beta with respect to S&P 500 index comes from the cash-flow components of the stocks added rather from the discount rates in both parts of the subsample. The changes in beta for the two subsamples are 0.230 and 0.533, estimated with t-stats above 3, where almost all the effect is cash-flow originated (0.297 and 0.393).

Secondly, I find that the difference in differences using matching stocks is also coming from the cash-flow components in both subsamples. Thirdly it is interesting to note that when breaking the sample in early and recent parts we observe that the change in beta related to discount rates is negative in the first part of the subsample and positive in the second part: -0.077 and 0.90 respectively significant at the 10% level of significance. This alone could be interpreted as evidence of sentiment-based comovement in the later part of the sample. However, we observe that the same pattern is observed in matched stocks, that were not added to the index (-0.061 and 0.084).

2.4.6.2 Subsample in growth value dimension In this subsection I study whether stocks with different characteristics differ in the change in beta experienced after inclusion. I divide the included firms into growth and value stocks, by comparing the cross-sectionally adjusted book-to-market ratios. Growth firms tend to be more intangible and more opaque, while value firms are more stable, if they are financially sound. Because the change in beta also reflects the size of the companies added, growth stocks should exhibit a higher change in beta than value stocks. Table 2.8 reports the results. Consistent with my prior, I find that the change in beta is higher for growth firms (0.547 versus 0.356). The results for matched firms exhibit similar patterns, and the difference in difference, although insignificant, is also coming from the cash-flow components of beta.

2.4.7 Robustness to a second-order VAR

After considering parsimonious VAR specifications, I turn now to test the results using richer VAR equations, both in the firm-level and in the aggregate. Recall that the news terms used in the benchmark event study around S&P 500 index inclusions are the sum of the news extracted from an aggregate VAR and a firm-level VAR. In the benchmark specification I only use one lag of the state variables, assuming that higher order lags would not affect present values of the variables, as widely used in the literature related to stock-return decomposition.

The benchmark aggregate specification assumes that the data generating process is a first-order monthly VAR. I use the following four state variables: excess return on the market (r^e_m), the term yield spread (TY), the log smoothed price-earnings ratio (PE), and the small-stock value spread (VS). Previous research has shown that these variables could help predict returns at a longer horizons (Campbell, Polk, and Vuolteenaho, 2010). Without being exhaustive (there are many possible specifications), I will test the results by using a second-order VAR, i.e., allowing for up to two lags to predict the state variables. The methodology is similar to the first order VAR:

$$z_{t+1} = a + A_1 z_t + A_2 z_{t-1} + u_{t+1} \tag{2.15}$$

which for analytical derivations of the news terms according to Campbell (1991), it can also be expressed as:

$$\begin{bmatrix} z_{t+1} \\ z_t \end{bmatrix} = \begin{bmatrix} a \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} z_t \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ 0 \end{bmatrix} \tag{2.16}$$

Table 2.9 shows the results for the second-order aggregate VAR. To avoid an unncessary display of zeros and the identity matrix, I only show A_1 and A_2. The results are similar to the first-order VAR. Due to the additional free parameters, however, the standard errors are somewhat larger. The coefficients for the second lag are estimated less accurately. Market returns exhibit now a bit of reversal in the second lag (with a coefficient of -0.04), term yield spread and price earning ratio keeps the positive sign in the second lag estimate, and the small stock value spread flips sign with respect to the first lag. The intercepts and the R-Squares are very similar to the previous specification.

I now turn to the firm-level market adjusted VAR. The variables used in the benchmark first-order VAR are the following: market adjusted log stock return (r_i), the previous year return, excluding the last month (MOM_i), and the log book-to-market (BM_i). I motivate this lag order as a second-order cointegrating VAR. Previous research has also shown that these variables have predictive power beyond the first month (Vuolteenaho, 2002, and Campbell, Polk, and Vuolteenaho, 2010). Consistent with Vuolteenaho (2002), I find that the results are very similar to the first-order VAR. Table 2.10 shows the coefficients of the secon-order market-adjusted firm-level VAR. As in the aggregate VAR, the standard errors of the second-order coefficients are large, and thus the coefficients are not accurately estimated. Monthly returns also exhibit reversal in the second lag, and the previous year return computed in the second lag predicts also positively the returns. The coefficient for book-to-market shows a different sign for the second lag, which is consistent with the first-order VAR given the degree of correlation between the book-to-market at time t and the book-to-market at time $t-1$.

Following the same methodology, I extract the news from each of the new VARs (the second-order aggregate VAR and second-order firm-level VAR), I add them up, yielding $N_{iDR,t}$ and $N_{iCF,t}$, and compute the changes in cash flow and discount rate betas after the addition in the S&P 500 index, as before. Table 2.11 shows the changes in overall beta (which I include again for comparison purposes), and the changes in the new cash flow and new discount rate betas. The main results are very robust to the use of a second order VAR. In column three we observe that the change in beta after an S&P 500 addition comes from the cash-flow beta. The overall change in beta is a stronly significant 0.430, the change in discount rate beta is an insignificant -0.035, and the change in cash flow beta is a stronly significant 0.424. Consistent with the results from the first order VAR, matched stocks also experience a change in the cash flow betas (column 6), and the difference in difference is all coming from the cash-flows (see column 9), although it is estimated less accurately.

2.4.8 An alternative specification of the VAR

In the benchmark specification, the cash-flow and discount-rate news are extracted from two different VARs. The rationale for estimating two different VARs hinged in the fact that firm-level idiosyncratic returns behave differently than market returns. A clear example shown in Tables 1 and 2 is that firm-level returns exhibit a clear short-term reversal after one month, while market returns display momentum after one month. Following Vuolteenaho (2002) and Campbell, Polk, and Vuolteenaho (2010), I estimated in the previous section an aggregate VAR to extract the market return news and a firm-level VAR to extract firm-level market adjusted returns, to account for the aforementioned differences and to more accurately predict the two components of a firm return: the idiosyncratic and the market component.

In this subsection I show that the main results of the paper are not driven by the choice of extracting the news from two different VARs. I now estimate a VAR for firm-level excess returns, instead of firm-level market-adjusted returns. In the state vector I now include firm-level and market-wide variables. By doing so, I intend to allow market-wide variables to affect expected returns and cash flows on all stocks. The model is then written this way:

$$\begin{bmatrix} z_{i,t+1} \\ x_{t+1} \end{bmatrix} = A + \Gamma \begin{bmatrix} z_{i,t} \\ x_t \end{bmatrix} + u_{i,t+1}$$

where $z_{i,t+1}$ is the vector of firm-specific variables, and the first element of this vector is the excess log return. Following Vuolteenaho (2002) I constrain the lower left corner of Γ to zero, which means that there is no feedback from firm-level state variables to market-wide state variables. Also, because the variables are not cross-sectionally demeaned, the do not necessarily have zero means, and thus and intercept vector A is included in the VAR.

Several specifications of the model are possible. In Table 2.12, I show the different options. This table only shows the first equation of the VAR for the different specifications (where the dependent variable is the firm-level excess log return). Firm-level variables include the excess log return, the previous year return (excluding the last month) in excess of the risk free rate during the same period, the log book to market ratio, and the log profitability in excess of the risk free rate. I include two sets of market-wide variables. The first one comprises the cross-sectional medians of the firm-level state variables, and the second one includes the four aggregate variables used to estimate the aggregate VAR in the previous section.

Columns (1) and (2) in Table 2.12 show the results when including the two different blocks of market-wide variables. In column (1) we can observe that all market-wide variables have predictive power consistent with previous literature, except the cross-sectional median of the variable MOM_t. In column (2) we also observe that all the aggregate variables have some predictive power as well, though not all them very significant. In order to have a relatively parsimonious VAR and choose the most significant variables, I conduct a horse-race of all the variables, as shown in column (3). Once all eight market-wide variables are included, we can see that three of the four cross-sectional medians cease to be significant, whereas the market return and term yield spread still have explanatory power. Although the cross-sectional median of profitability is significant in this specification, it appears insignificant if the insignificant variables are dropped (this and other horse-race options have been evaluated but not shown for the sake of brevity). The final set of variables I use are the ones shown in column (4).

Table 2.13 shows all the coefficients for the VAR corresponding to column (4) in the previous table. Intercepts are included in the VAR, however the magnitude is very small and insignificant in all cases. All state variables in the first equation are significant at the 1%. The sign of the variables is as expected: the coefficient for excess log return is negative (showing the short-term reversal), positive and strong for momentum, profitability, market return and term yield spread. The equations corresponding to the aggregate variables are consistent with the aggregate VAR estimated in the previous section: market return exhibits momentum at the monthly level, and term yield spread predicts positively market return. The R-Square, 2%, is also similar (although lower, because there are only two variables predicting market returns now) to the previous aggregate VAR 2.81%.

I then extract the news from this new VAR, $N_{iDR,t}$ and $N_{iCF,t}$, and compute the changes in cash flow and discount rate betas after the addition in the S&P 500 index, as before. The only difference is that I now estimate the betas with different news, the ones extracted from this alternative specification of the VAR. Table 2.14 shows the changes in overall beta (included again for comparison purposes), in the new cash-flow and discount rate betas. The main results are robust to this different specification of the VAR. In column three we observe that the change in beta after an S&P 500 addition comes from the cash-flow beta. The overall change in beta is a stronly significant 0.430, the change in discount rate beta is an insignificant 0.036, and the change in cash flow beta is a stronly significant 0.357. And as in the previous Section, when compared the changes in betas with matched stocks, the difference in difference is all coming from the cash-flows, and is less significant than for the event stocks.

2.4.9 Alternative cash flow risk measure

There is a recent novel method of estimating cash-flow news alternative to the use of a VAR decomposition, suggested by Da and Warachka (2009). They use revisions in analyst earnings forecasts to construct an analyst earnings beta, that measures the covariance between the cash flow innovations of a stock and those of the market. Empirical analysis of S&P 500 index inclusions using this specification yields results more ambiguous than the ones derived from the VAR procedure. This is not surprising, as Da and Warachka (2009) also show that their results are not consistent with the use of cash-flow news extracted from a VAR.

2.5 Conclusion

Using a two beta decomposition, I provide evidence of changes in cash-flow covariances after stock additions to the S&P 500 index. I show that the well-known beta change effect after index inclusions is associated with the cash-flow news components of the individual stocks that are added into the index. These results are robust to alternative specifications of the VAR, such a second-order VAR, and a unique VAR that encompasses firm-level and aggregate variables as state variables.

I also study direct measures of cash flows, coming from accounting variables, as a robustness check of my VAR approach, and show that the results do not depend on my particular specification.

The results from the benchmark study, from a matching procedure and from subsample analysis, as well as from a direct approach, are consistent with a story where it is Standard and Poors *selecting* stocks that will exhibit a growth in betas.

2.6 Tables

Table 2.1: Aggregate VAR

This table shows the OLS parameter estimates for a first-order monthly aggregate VAR model including a constant, the log excess market return (r_M^e), 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.

Aggregate VAR to predict market return

	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%
(Log excess market return)	(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%
(Term yield spread)	(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%
(Log price-earnings ratio)	(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%
(Small-stock value spread)	(0.0169)	(0.0283)	(0.0022)	(0.0043)	(0.0048)	

Table 2.2: Firm-level VAR

This 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). I 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.

Firm-level VAR for market-adjusted returns

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

Table 2.3: Changes in Beta - Benchmark Case

This table shows the changes in the slope of regressions of returns of stocks added to the S&P 500 on returns of the S&P 500 index and the non-S&P 500 rest of the market. The sample includes those stocks added to the S&P 500 between 1976 and 2008 that were not involved in mergers or related events around the stock addition. I estimate a pooled regression with data from 36 months before to 36 months after the addition. I interact the returns on the S&P 500 and the non S&P 500 with a dummy I_{it} that takes value 1 if the stock is in the index. This way, the coefficient associated with the interaction terms reveals the change in beta after the addition. The bivariate regression estimated is the following:

$$r^e_{i,t} = \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t}$$

The excess return on the S&P 500 index, r^e_{SP}, is computed as the difference between the monthly return on the S&P 500 index, obtained from the CRSP Index File, and the monthly riskfree rate, obtained from Professor Kenneth French's website. The return r^e_{nSP} are excess returns on a capitalization-weighted index of the non-S&P 500 stocks in the NYSE, AMEX, and Nasdaq, and are inferred from the following identity:

$$r_{M,t} = \left(\frac{CAP_{M,t-1} - CAP_{SP,t-1}}{CAP_{M,t-1}} \right) r_{nSP,t} + \left(\frac{CAP_{SP,t-1}}{CAP_{M,t-1}} \right) r_{SP,t}$$

where total capitalization of the S&P 500 (CAP_{SP}) is from the CRSP Index on the S&P 500 Universe file. Returns on the value-weighted CRSP NYSE, AMEX, and Nasdaq index (r_M) and total capitalization (CAP_M) are from the CRSP Stock Index file. I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

	$r^e_{i,t}$
$r^e_{SP,t}$	0.550***
	(0.082)
$I_{it} r^e_{SP,t}$	0.425***
	(0.068)
$r^e_{nSP,t}$	0.557***
	(0.067)
$I_{it} r^e_{nSP,t}$	-0.291***
	(0.062)
Constant	0.007***
	(0.001)
Observations	24016
R-squared	0.253

Table 2.4: Changes in cash-flow and discount rate betas

This table shows the changes in the slope of regressions of returns (and its components) of stocks added to the S&P 500 on returns of the S&P 500 index and the non-S&P 500 rest of the market. The sample and definition of variables is described in Table 2.3. This table shows the results of regressions similar to the previous table, but replacing the returns on the left hand side variable with (negative of) discount-rate news ($-N_{i,DR}$) and cash-flow news ($N_{i,CF}$) of the event stocks. The equations estimated are the following:

$$\begin{aligned}
r^e_{i,t} &= \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\
-N_{iDR,t} &= \alpha_i + \beta^{DRb}_{SP} r^e_{SP,t} + \beta^{DRb}_{nSP} r^e_{nSP,t} + \Delta\beta^{DR}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{DR}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\
N_{iCF,t} &= \alpha_i + \beta^{CFb}_{SP} r^e_{SP,t} + \beta^{CFb}_{nSP} r^e_{nSP,t} + \Delta\beta^{CF}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{CF}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t}
\end{aligned}$$

I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.550***	0.629***	-0.107
	(0.082)	(0.065)	(0.108)
$I_{it} r^e_{SP,t}$	0.425***	-0.008	0.391***
	(0.068)	(0.036)	(0.059)
$r^e_{nSP,t}$	0.557***	0.249***	0.209**
	(0.067)	(0.056)	(0.087)
$I_{it} r^e_{nSP,t}$	-0.291***	0.049*	-0.286***
	(0.062)	(0.029)	(0.057)
Constant	0.007***	-0.001	0.001
	(0.001)	(0.001)	(0.002)
Observations	24016	24016	24016
R-squared	0.253	0.607	0.024

Table 2.5: Direct measures of cash flows

This table shows the changes in the slope of regressions of return on equity of stocks added to the S&P 500 on return on equity of the S&P 500 index and the return on equity of non-S&P 500 rest of the market. The sample includes those stocks added to the S&P 500 between 1976 and 2008 that were not involved in mergers or related events around the stock addition. I interact the returns on the S&P 500 and the non S&P 500 with a dummy I_{it} that takes value 1 if the stock is in the index. This way, the coefficient associated with the interaction terms reveals the change in beta after the addition. The equation I estimate is:

$$roe_{i,t} = \alpha_i + \beta^b_{SP} roe_{SP,t} + \beta^b_{nSP} roe_{nSP,t} + \Delta\beta_{SP} I_{it} roe_{SP,t} + \Delta\beta_{nSP} I_{it} roe_{nSP,t} + \varepsilon_{i,t}$$

where roe_{it} is the log of return on equity, defined as $roe_{it} = \log(1+NI_t/BE_{t-1})$ where NI is net income and BE book equity, in t and $t-1$ respectively. To avoid extreme observations, ROE_{it} is winsorized between -1 and 3 (on a given quarter, the return on equity cannot be lower than -100% or higher than 300%). $roe_{SP,t}$ and $roe_{nSP,t}$ are calculated as the log of 1 plus the sum of NI_t over the sum of BE_{t-1}, for all December fiscal year end stocks in each group of S&P 500 and non S&P 500 stocks. As in the previous analyses, I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

	$roe_{i,t}$
$roe_{SP,t}$	0.227***
	(0.080)
$I_{it}roe_{SP,t}$	0.261**
	(0.122)
$roe_{nSP,t}$	0.716***
	(0.106)
$I_{it}roe_{nSP,t}$	-0.505***
	(0.150)
Constant	0.011***
	(0.003)
R-squared	0.170

Table 2.6: Changes in beta and matching firms

This table shows the changes in the slope of regressions of returns (and its components) of stocks added to the S&P 500, matched stocks, and their difference, on returns of the S&P 500 index and the non-S&P 500 rest of the market. Firms are matched to event stocks based on industry and size, as described in the text. The sample and definition of variables is described in Table 2.3. The equations estimated are the following:

$$\begin{aligned} r^e_{i,t} &= \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\ -N_{iDR,t} &= \alpha_i + \beta^{DRb}_{SP} r^e_{SP,t} + \beta^{DRb}_{nSP} r^e_{nSP,t} + \Delta\beta^{DR}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{DR}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\ N_{iCF,t} &= \alpha_i + \beta^{CFb}_{SP} r^e_{SP,t} + \beta^{CFb}_{nSP} r^e_{nSP,t} + \Delta\beta^{CF}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{CF}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \end{aligned}$$

I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

	Event Firms			Matched Firms			Difference		
	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.539***	0.628***	-0.114	0.623***	0.637***	-0.046	-0.081	-0.009	-0.065
	(0.080)	(0.064)	(0.106)	(0.067)	(0.063)	(0.093)	(0.053)	(0.017)	(0.053)
$I_{it} r^e_{SP,t}$	0.430***	-0.005	0.392***	0.261***	0.002	0.230***	0.165*	-0.006	0.158**
	(0.068)	(0.035)	(0.061)	(0.085)	(0.035)	(0.079)	(0.086)	(0.024)	(0.079)
$r^e_{nSP,t}$	0.555***	0.249***	0.203**	0.411***	0.250***	0.084	0.142***	-0.002	0.118***
	(0.066)	(0.056)	(0.087)	(0.060)	(0.055)	(0.080)	(0.044)	(0.014)	(0.042)
$I_{it} r^e_{nSP,t}$	-0.298***	0.043	-0.284***	-0.177**	0.036	-0.176**	-0.120*	0.006	-0.106*
	(0.060)	(0.029)	(0.053)	(0.076)	(0.028)	(0.072)	(0.069)	(0.020)	(0.062)
Constant	0.007***	-0.001	0.001	0.003***	-0.001	-0.001	0.003***	-0.000	0.003***
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.000)	(0.001)
Observations	21118	21118	21118	21066	21066	21066	21066	21066	21066
R-squared	0.249	0.610	0.023	0.234	0.614	0.012	0.013	0.005	0.011

Table 2.7: Robustness to time subsamples

This table follows the same sample and procedure as table 2.6. Panel A shows the results using only the bottom quintile of event stocks sorted on book-to-market ratios. Panel B shows the results using only the top quintile of event stocks sorted on book-to-market ratios. The equations estimated are similar to those in table 2.6. I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

1976-1991	Event Firms			Matched Firms			Difference		
	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.375***	0.793***	-0.420***	0.488***	0.772***	-0.297***	-0.106	0.022	-0.119*
	(0.065)	(0.072)	(0.085)	(0.062)	(0.067)	(0.080)	(0.071)	(0.023)	(0.067)
$I_{it}r^e_{SP,t}$	0.230***	-0.077*	0.297***	-0.043	-0.061	0.038	0.266***	-0.016	0.253***
	(0.067)	(0.045)	(0.066)	(0.071)	(0.043)	(0.069)	(0.096)	(0.026)	(0.080)
$r^e_{nSP,t}$	0.733***	0.106*	0.501***	0.578***	0.125**	0.354***	0.150**	-0.021	0.143**
	(0.054)	(0.061)	(0.076)	(0.054)	(0.057)	(0.077)	(0.060)	(0.020)	(0.055)
$I_{it}r^e_{nSP,t}$	-0.150**	0.067	-0.194***	0.036	0.057	-0.029	-0.180**	0.011	-0.160**
	(0.063)	(0.041)	(0.061)	(0.066)	(0.039)	(0.064)	(0.083)	(0.022)	(0.067)
Observations	11482	11482	11482	11463	11463	11463	11463	11463	11463
R-squared	0.339	0.663	0.042	0.324	0.670	0.028	0.014	0.004	0.014

1991-2005	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.584***	0.534***	0.016	0.632***	0.568***	0.028	-0.049	-0.035	-0.011
	(0.121)	(0.084)	(0.152)	(0.101)	(0.088)	(0.137)	(0.074)	(0.022)	(0.073)
$I_{it}r^e_{SP,t}$	0.533***	0.090*	0.393***	0.448***	0.084	0.314***	0.084	0.007	0.075
	(0.111)	(0.049)	(0.097)	(0.113)	(0.060)	(0.110)	(0.129)	(0.036)	(0.118)
$r^e_{nSP,t}$	0.477***	0.319***	0.062	0.343***	0.313***	-0.040	0.135**	0.007	0.103*
	(0.091)	(0.070)	(0.108)	(0.084)	(0.074)	(0.107)	(0.060)	(0.017)	(0.056)
$I_{it}r^e_{nSP,t}$	-0.325***	0.002	-0.260***	-0.234**	0.001	-0.182**	-0.092	0.001	-0.078
	(0.087)	(0.036)	(0.075)	(0.090)	(0.042)	(0.089)	(0.089)	(0.027)	(0.082)
Observations	9636	9636	9636	9603	9603	9603	9603	9603	9603
R-squared	0.188	0.556	0.018	0.175	0.559	0.012	0.012	0.005	0.010

Table 2.8: Robustness to growth-value subsamples

This table follows the same sample and procedure as table 2.6. Panel A shows the results using only the bottom quintile of event stocks sorted on book-to-market ratios. Panel B shows the results using only the top quintile of event stocks sorted on book-to-market ratios. The equations estimated are similar to those in table 2.6. I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

LOW BM	Event Firms			Matched Firms			Difference		
	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.494***	0.528***	-0.064	0.589***	0.577***	-0.009	-0.094	-0.051**	-0.051
	(0.125)	(0.076)	(0.145)	(0.128)	(0.079)	(0.137)	(0.090)	(0.024)	(0.087)
$I_{it}r^e_{SP,t}$	0.547***	0.049	0.447***	0.333**	0.046	0.255*	0.211	0.007	0.185
	(0.133)	(0.060)	(0.121)	(0.154)	(0.062)	(0.134)	(0.147)	(0.039)	(0.132)
$r^e_{nSP,t}$	0.715***	0.313***	0.286**	0.560***	0.302***	0.171	0.156**	0.012	0.115*
	(0.103)	(0.064)	(0.112)	(0.107)	(0.066)	(0.108)	(0.077)	(0.018)	(0.070)
$I_{it}r^e_{nSP,t}$	-0.300***	0.002	-0.244**	-0.218*	0.006	-0.202*	-0.081	-0.006	-0.039
	(0.105)	(0.047)	(0.100)	(0.129)	(0.047)	(0.117)	(0.115)	(0.029)	(0.102)
Observations	7459	7459	7459	7432	7432	7432	7432	7432	7432
R-squared	0.249	0.559	0.038	0.222	0.576	0.018	0.013	0.007	0.013

HIGH BM	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.548***	0.730***	-0.200*	0.635***	0.705***	-0.109	-0.082	0.026	-0.088
	(0.089)	(0.058)	(0.107)	(0.088)	(0.058)	(0.102)	(0.080)	(0.019)	(0.065)
$I_{it}r^e_{SP,t}$	0.356***	-0.037	0.355***	0.223***	-0.015	0.218***	0.128	-0.022	0.132
	(0.092)	(0.038)	(0.084)	(0.085)	(0.037)	(0.074)	(0.103)	(0.026)	(0.084)
$r^e_{nSP,t}$	0.470***	0.172***	0.201**	0.320***	0.184***	0.066	0.145**	-0.014	0.132**
	(0.074)	(0.049)	(0.092)	(0.081)	(0.049)	(0.094)	(0.071)	(0.016)	(0.055)
$I_{it}r^e_{nSP,t}$	-0.282***	0.074**	-0.310***	-0.173***	0.051*	-0.191***	-0.103	0.023	-0.115
	(0.078)	(0.032)	(0.071)	(0.066)	(0.030)	(0.059)	(0.086)	(0.021)	(0.071)
Observations	9355	9355	9355	9341	9341	9341	9341	9341	9341
R-squared	0.254	0.635	0.016	0.250	0.630	0.013	0.013	0.004	0.010

Table 2.9: Second Order Aggregate VAR

This table shows the OLS parameter estimates for a second-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 nine columns report coefficients on the nine explanatory variables and the tenth column reports the corresponding adjusted R^2. Standard errors are in parentheses. The sample period for the dependent variables is January 1929 - May 2009, providing 965 monthly data points.

Firm-level VAR for excess returns

Variable	*Intercept*	$r^e_{M,t}$	TY_t	PE_t	VS_t	$r^e_{M,t-1}$	TY_{t-1}	PE_{t-1}	VS_{t-1}	R^2
$r^e_{M,t+1}$	0.0672	0.0929	0.0014	0.0235	-0.0328	-0.0408	0.0029	-0.0398	0.0211	0.0294
(Log excess market return)	(0.0191)	(0.0513)	(0.0065)	(0.0770)	(0.0364)	(0.0478)	(0.0065)	(0.0772)	(0.0365)	
TY_{t+1}	-0.0160	-0.2856	0.7763	0.5627	-0.0035	-0.5058	0.1584	-0.5686	0.0563	0.8679
(Term yield spread)	(0.0938)	(0.2520)	(0.0319)	(0.3783)	(0.1790)	(0.2351)	(0.0319)	(0.3794)	(0.1793)	
PE_{t+1}	0.0121	0.7585	0.0000	0.5379	-0.0359	0.1720	0.0015	0.4571	0.0342	0.9918
(Log price-earnings ratio)	(0.0122)	(0.0328)	(0.0042)	(0.0492)	(0.0233)	(0.0306)	(0.0042)	(0.0493)	(0.0233)	
VS_{t+1}	0.0235	-0.0967	0.0093	0.1916	1.0240	-0.0647	-0.0092	-0.1938	-0.0341	0.9827
(Small-stock value spread)	(0.0169)	(0.0454)	(0.0057)	(0.0681)	(0.0322)	(0.0423)	(0.0057)	(0.0683)	(0.0323)	

Table 2.10: Second Order Firm-level VAR

This table shows the pooled-WLS parameter estimates for a second-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). I 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.

Second order firm-level VAR for market-adjusted returns

Variable	$r_{i,t}$	$MOM_{i,t}$	$BM_{i,t}$	$r_{i,t-1}$	$MOM_{i,t-1}$	$BM_{i,t-1}$	R^2
$r_{i,t+1}$	-0.0336	0.0189	0.0210	-0.0109	0.0028	-0.0162	0.0066
(Log return)	(0.0085)	(0.0054)	(0.0036)	(0.0082)	(0.0052)	(0.0035)	
$MOM_{i,t+1}$	0.9584	0.8365	-0.0025	0.1551	0.0593	0.0018	0.9197
(1Y Momentum)	(0.0070)	(0.0065)	(0.0032)	(0.0074)	(0.0057)	(0.0033)	
$BM_{i,t+1}$	0.0270	-0.0261	0.9631	0.0211	0.0151	0.0235	0.9709
(Log BM)	(0.0053)	(0.0036)	(0.0041)	(0.0054)	(0.0035)	(0.0038)	

Table 2.11: Changes in beta and matching firms for a second-order VAR

This table shows the changes in the slope of regressions of returns (and its components extracted from a second-order VAR) of stocks added to the S&P 500, matched stocks, and their difference, on returns of the S&P 500 index and the non-S&P 500 rest of the market. Firms are matched to event stocks based on industry and size, as described in the text. The sample and definition of variables is described in Table 2.3. The equations estimated are the following:

$$\begin{aligned}
r^e_{i,t} &= \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\
-N_{iDR,t} &= \alpha_i + \beta^{DRb}_{SP} r^e_{SP,t} + \beta^{DRb}_{nSP} r^e_{nSP,t} + \Delta\beta^{DR}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{DR}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\
N_{iCF,t} &= \alpha_i + \beta^{CFb}_{SP} r^e_{SP,t} + \beta^{CFb}_{nSP} r^e_{nSP,t} + \Delta\beta^{CF}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{CF}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t}
\end{aligned}$$

I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

	Event Firms			Matched Firms			Difference		
	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.539***	0.652***	-0.140	0.625***	0.660***	-0.069	-0.084	-0.032	-0.044
	(0.080)	(0.074)	(0.120)	(0.067)	(0.072)	(0.105)	(0.054)	(0.033)	(0.057)
$I_{it} r^e_{SP,t}$	0.430***	-0.035	0.424***	0.261***	-0.025	0.259***	0.165*	0.021	0.129
	(0.068)	(0.037)	(0.061)	(0.085)	(0.037)	(0.082)	(0.086)	(0.025)	(0.083)
$r^e_{nSP,t}$	0.555***	0.336***	0.118	0.411***	0.340***	-0.004	0.142***	-0.092***	0.205***
	(0.066)	(0.064)	(0.101)	(0.060)	(0.063)	(0.094)	(0.044)	(0.027)	(0.047)
$I_{it} r^e_{nSP,t}$	-0.298***	0.061**	-0.304***	-0.178**	0.052*	-0.194***	-0.119*	-0.009	-0.086
	(0.060)	(0.030)	(0.055)	(0.076)	(0.029)	(0.075)	(0.069)	(0.021)	(0.065)
Constant	0.007***	-0.004***	0.004**	0.003***	-0.004***	0.001	0.004***	0.002***	0.000
	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.002)	(0.001)	(0.001)	(0.001)
Observations	21118	21114	21114	21066	21051	21051	21066	21051	21051
R-squared	0.249	0.648	0.018	0.235	0.653	0.013	0.013	0.034	0.018

Table 2.12: Alternative VAR: different specifications

This table shows the pooled-WLS parameter estimates for the first equation of a first-order monthly firm-level VAR model. The state variables include a constant, a set of firm-level variables, and two sets of aggregate variables. The firm level variables are: the excess log stock return ($r^e_{i,t}$), stock momentum ($MOM^e_{i,t}$), the log book-to-market ratio ($BM_{i,t}$), and the log profitability in excess of the risk free rate. The first set of aggregate variables is formed by the cross-sectional median of each of the firm-level variables. The second set of aggregate variables consists of 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). Standard errors are in parentheses. 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.

Predicting firm-level excess returns, dependent variable: $r^e_{i,t+1}$

	(1)	(2)	(3)	(4)
Constant	0.0000	0.0000	0.0000	0.0000
	(0.0000)	(0.0000)	(0.0000)	(0.0000)
$r^e_{i,t}$	-0.0526***	-0.0478***	-0.0524***	-0.0477***
	(0.0060)	(0.0075)	(0.0060)	(0.0076)
$MOM^e_{i,t}$	0.0170***	0.0146***	0.0171***	0.0151***
	(0.0023)	(0.0046)	(0.0022)	(0.0045)
$BM_{i,t}$	0.0050***	0.0069***	0.0050***	0.0073***
	(0.0007)	(0.0014)	(0.0007)	(0.0015)
$ROE^e_{i,t}$	0.0135***	0.0184***	0.0141***	0.0206***
	(0.0022)	(0.0025)	(0.0021)	(0.0035)
median $r^e_{i,t}$	0.2950***		0.0928	
	(0.0488)		(0.1041)	
median $MOM^e_{i,t}$	-0.0072		-0.0162	
	(0.0149)		(0.0158)	
median $BM_{i,t}$	0.0240***		0.0147	
	(0.0099)		(0.0142)	
median $ROE^e_{i,t}$	0.1534***		0.2742**	
	(0.0522)		(0.1362)	
$r^e_{M,t}$		0.2698***	0.1926*	0.2753***
		(0.0490)	(0.1037)	(0.0479)
TY_t		0.0059*	0.0050*	0.0068**
		(0.0030)	(0.0031)	(0.0031)
PE_t		-0.0103*	-0.0116	
		(0.0059)	(0.0102)	
VS_t		0.0224*	0.0146	
		(0.0119)	(0.0153)	
$\bar{R}^2$	0.0203	0.0192	0.0226	0.0182
Observations	1,658,049	1,658,049	1,658,049	1,658,049

Table 2.13: Alternative VAR: predicting firm-level excess returns

This table shows the pooled-WLS parameter estimates for all the equations in the first-order monthly firm-level VAR model corresponding to column (4) in the previous Table. The model state vector includes a set of firm-level variables: the excess log stock return ($r^e_{i,t}$), stock momentum ($MOM^e_{i,t}$), the log book-to-market ratio ($BM_{i,t}$), and the log profitability in excess of the risk free rate; and a set of aggregate variables: log excess market return (r^e_M), and the term yield spread (TY). Rows corresponds to dependent variables and columns to independent (lagged dependent) variables. The first seven columns report coefficients on the seven explanatory variables and the eighth 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.

Firm-level VAR for excess returns

Variable	*Intercept*	$r_{i,t}$	$MOM_{i,t}$	$BM_{i,t}$	$ROE_{i,t}$	$r^e_{M,t}$	TY_t	R^2
$r_{i,t+1}$	0.0000	-0.0477	0.0151	0.0073	0.0206	0.2753	0.0068	0.0182
(Log stock return)	(0.0000)	(0.0076)	(0.0045)	(0.0015)	(0.0035)	(0.0479)	(0.0031)	
$MOM_{i,t+1}$	0.0000	0.9540	0.9106	-0.0009	-0.0104	0.0264	0.0060	0.9237
(One year momentum)	(0.0000)	(0.0066)	(0.0040)	(0.0015)	(0.0031)	(0.0432)	(0.0027)	
$BM_{i,t+1}$	0.0000	0.0432	-0.0120	0.9842	-0.0033	-0.2620	-0.0055	0.9701
(Log book-to-market)	(0.0000)	(0.0075)	(0.0044)	(0.0017)	(0.0037)	(0.0511)	(0.0031)	
$ROE^e_{i,t+1}$	0.0000	0.0030	0.0111	-0.0042	0.9511	0.0024	0.0004	0.8980
(Profitability)	(0.0000)	(0.0012)	(0.0009)	(0.0005)	(0.0057)	(0.0048)	(0.0003)	
$r^e_{M,t+1}$	0.0000	0	0	0	0	0.1110	0.0056	0.0200
(Log excess market return)	(0.0000)					(0.0491)	(0.0029)	
TY_{t+1}	0.0000	0	0	0	0	0.3655	0.8865	0.7886
(Term yield spread)	(0.0000)					(0.3420)	(0.0203)	

Table 2.14: Changes in beta and matching firms for an alternative VAR

This table shows the changes in the slope of regressions of returns (and its components) of stocks added to the S&P 500, matched stocks, and their difference, on returns of the S&P 500 index and the non-S&P 500 rest of the market. Firms are matched to event stocks based on industry and size, as described in the text. The sample and definition of variables is described in Table 2.3. The equations estimated are the following:

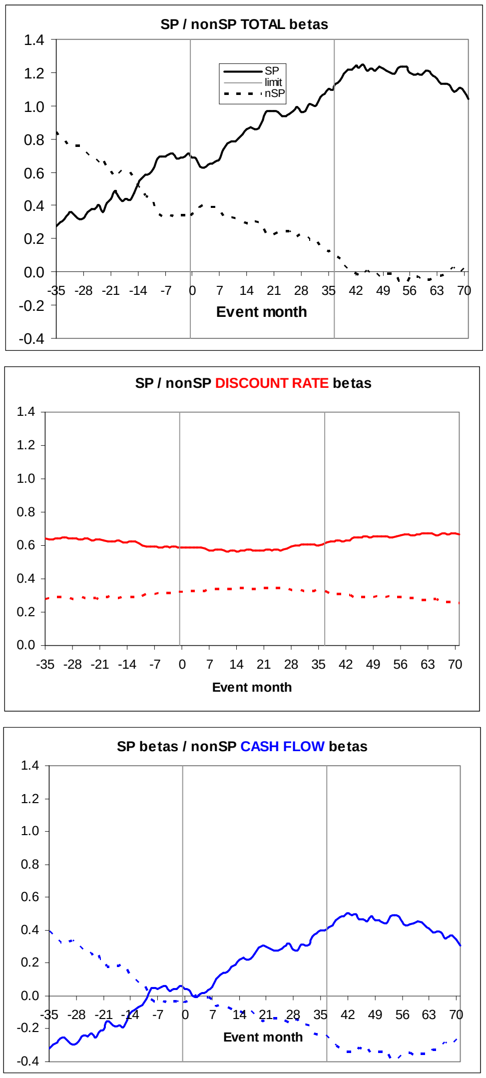
$$
\begin{aligned}
r^e_{i,t} &= \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\
-N_{iDR,t} &= \alpha_i + \beta^{DRb}_{SP} r^e_{SP,t} + \beta^{DRb}_{nSP} r^e_{nSP,t} + \Delta\beta^{DR}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{DR}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \\
N_{iCF,t} &= \alpha_i + \beta^{CFb}_{SP} r^e_{SP,t} + \beta^{CFb}_{nSP} r^e_{nSP,t} + \Delta\beta^{CF}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{CF}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t}
\end{aligned}
$$

I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

	Event Firms			Matched Firms			Difference		
	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.539***	-0.090	0.620***	0.625***	-0.080	0.692***	-0.084	-0.010	-0.069
	(0.080)	(0.078)	(0.102)	(0.067)	(0.075)	(0.107)	(0.054)	(0.019)	(0.053)
$I_{it} r^e_{SP,t}$	0.430***	0.036	0.357***	0.261***	0.055	0.181**	0.165*	-0.017	0.171**
	(0.068)	(0.030)	(0.067)	(0.085)	(0.036)	(0.075)	(0.086)	(0.029)	(0.081)
$r^e_{nSP,t}$	0.555***	-0.132**	0.580***	0.411***	-0.132**	0.460***	0.142***	0.001	0.118***
	(0.066)	(0.057)	(0.070)	(0.060)	(0.053)	(0.071)	(0.044)	(0.016)	(0.043)
$I_{it} r^e_{nSP,t}$	-0.298***	-0.004	-0.256***	-0.178**	-0.025	-0.131*	-0.119*	0.019	-0.122*
	(0.060)	(0.025)	(0.053)	(0.076)	(0.030)	(0.068)	(0.069)	(0.024)	(0.064)
Constant	0.007***	-0.001	-0.000	0.003***	-0.001	-0.003	0.004***	0.000	0.003***
	(0.001)	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)	(0.001)	(0.000)	(0.001)
Observations	21118	21118	21118	21066	21066	21066	21066	21066	21066
R-squared	0.249	0.064	0.312	0.235	0.059	0.301	0.013	0.006	0.012

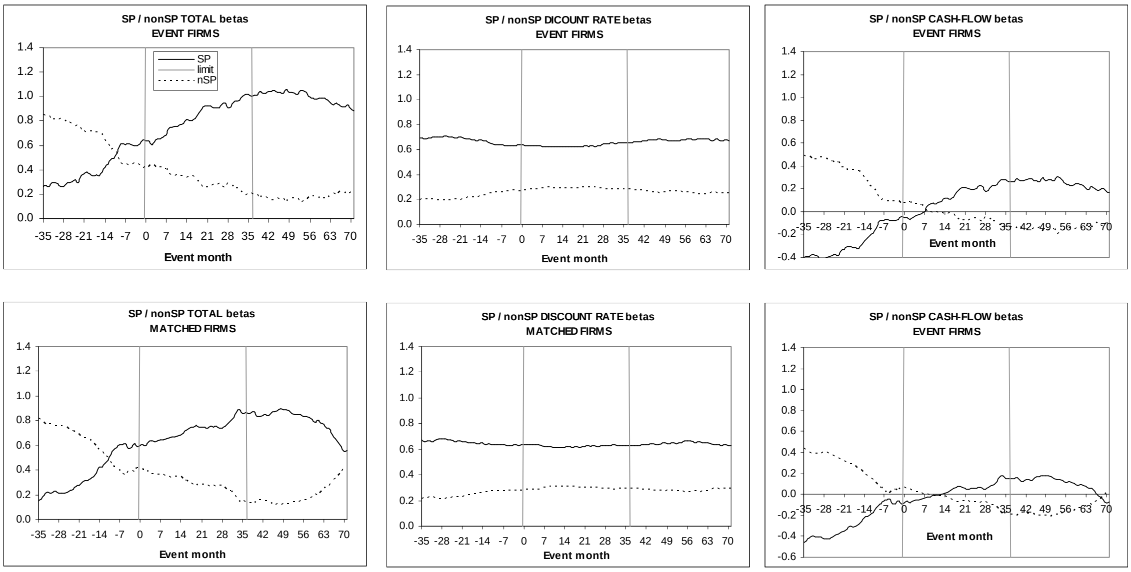
2.7 Figures

Figure 2.1: Evolution of rolling betas around S&P 500 index inclusions



This figure plots the evolutions of rolling betas around S&P 500 index inclusions. In the top panel I plot the evolution of the overall beta, in the mid panel I show the evolution of discount-rate betas, and in the bottom panel the evolution of cash-flow betas.

Figure 2.2: Evolution of rolling betas around S&P 500 index inclusions for event and matched stocks



This figure shows rolling betas around S&P 500 index inclusions. From left to right it shows total, discount rate, and cash-flow betas. The top panel shows the evolution of betas for event stocks, and the bottom panel shows the evolution of betas for matched stocks.

3 Comovement in the CDS Market

(Joint work with Daniel Bergstresser)

3.1 Introduction

Economic theory suggests that in a frictionless economy with rational investors, securities' prices should at all times reflect their fundamental values. In this idealized setting, comovement in the securities' values and returns should reflect only comovement in underlying fundamentals. Recent research, however, documents comovement in securities' returns that appears to exceed fundamental comovement. This research includes work on US equity markets by Barberis, Shleifer, and Wurgler (2005), analysis of Japanese equity markets by Greenwood and Sosner (2007), and earlier work by Vijh (1994).

Research on comovement in equity markets has often used inclusion in and deletion from benchmark indexes as part of the research design. Many mutual funds and exchange traded funds are explicitly tied to these benchmark indexes. The flow of investors' money into and out of these funds induces correlation in trading activity across the index constituents. In a frictionless market this correlated trading would have no effect on prices or returns. But frictions and illiquidity, even among relatively liquid equity securities, appear to induce excessive index-based comovement in American and international equity markets.

This paper extends the existing literature by exploring index-based comovement in the market for Credit Default Swaps (CDS). CDS contracts are derivative contracts whose cashflows are tied to credit events at underlying bond issuers. An investor who has sold protection on an issuer using a CDS contract has taken on that issuers' credit risk, similar to the purchaser of the issuers' bonds. Like equity markets, CDS markets have several benchmark indexes. These indexes are used both as barometers for market activity and as trading instruments in their own right. We use the most liquid CDS index benchmark: the Markit North American Investment Grade CDX index (CDX.NA.IG hereafter). The index's constituents are updated biannually, providing a large sample of inclusion and deletion events for our analysis.

Because bonds and CDS contracts both offer investors economic exposure to an issuers' credit risk, exploring comovement in the two markets jointly allows us to control for fundamentals-based comovement. This approach for controlling for underlying fundamentals has not been available to researchers analyzing comovement in equity markets. With index inclusions, we find that comovement of CDS spreads with the other issuers in the index increases significantly around the inclusion date CDX. The mean beta against the index rises 0.284 after inclusion. The difference in differences of mean betas from CDS spreads and bonds is a statistically significant 0.301 after inclusion. This evidence supports the hypothesis that the bond and CDS markets are at least somewhat segmented. Index inclusion appears to change the comovement patterns of CDS spreads in a way that is not matched by the comovement patterns of the underlying bonds.

To better understand the source of this non-fundamental comovement, we also estimate Dimson (1979) betas. We find that our results are very strong even using Dimson betas, which suggests that the origin of this shift in comovement is not an information diffusion channel, but rather a category based explanation for non-fundamental comovement. Many investors buy protection in baskets, buy the index, however they do not buy individual CDS. This clientele effect is translated into an excess comovement of those CDS that are part of the index.

Though most of our analysis is focused on additions to the index, we also show that deletions from the index see no statistically significant change in the mean beta of the CDS on the index. The betas are high prior to deletions because issuers being deleted from the CDX Investment Grade index are often being removed because they lose their investment grade status: as firms approach distress their bonds begin to take on a larger share of the company's risk. On net, these results indicate that index-based comovement is a characteristic of CDS markets as well as equity markets.

The paper proceeds in five sections. Sections I and II review in more detail the relevant literatures on comovement and on CDS markets. Section III and IV describe the empirical design and the data used in the study. Results are presented and discussed in Section V. A brief final section concludes.

3.2 Related literature on comovement

A number of researchers have investigated patterns of comovement in equity prices. Research has focused on whether patterns of comovement reflect joint movement in expected returns and rational discount rates, or rather are driven by commonality of trading activity across different securities. Pindyck and Rotemberg (1993) focus on US equity securities, estimate a factor model of stock price returns similar to Chen, Roll, and Ross (1986), and find comovement across the residuals from this regression. They show that comovement is particularly large among stocks held by institutional investors, which they interpret as indicating that these investors' flows drive securities away from fundamental value.

Vijh (1994) looked at the betas of securities included and excluded from the S&P 500, showing that securities in the S&P 500 have higher betas. Vijh estimates that 8.5 percent of the total variance of daily returns of the market portfolio is based on flow-related price pressure. Barberis, Shleifer, and Wurgler (2005) also focus on the S&P 500 index inclusions and deletions and find evidence of comovement in excess of what can be explained with fundamentals.

Greenwood (2005) focuses on Japan, and exploits the fact that the Nikkei 225 index is equally weighted, rather than value-weighted. Some stocks in the index are thus overweighted by a factor of ten or more relative to other stocks in the index. Thus, when investor demand for the Nikkei 225 index rises, investors have to purchase significantly more of some stocks (relative to value) than they would if the index were value-weighted. In particular, firms with small market capitalizations have larger demand shocks, relative to size. Greenwood and Sosner (2007) also focus on Japan, on the April 2000 redefinition of the Nikkei 225 index. Daily index return betas of the additions rose by an average of 0.45; index return betas of the deleted stocks fell by an average of 0.63.

Antón and Polk (2009) have investigated comovement in a bottom-up framework, and find that stocks that are held by the same active fund managers and covered by the same analysts comove more than other stocks, controlling for other similarities between stocks. This effect is stronger when the stocks in the pair are small and common owners are experiencing strong inflows and outflows. A related paper by Greenwood and Thesmar (2009) develops and applies a measure of 'co-fragility' in US equity markets, that captures the correlation of the trading needs of two assets' owners: two assets are 'co-fragile' if they are held by investors with correlated inflows and outflows. Another related paper by Koch, Ruenzi, and Starks (2009) looks at comovement among stocks with high and low institutional ownership, and find that the stocks with high mutual fund ownership have comovement that is twice as pronounced as among stocks with minimal institutional ownership.

Evans and Lyons (2002) investigate trading-based price pressure in the currency market, and find that order flow explains a very significant share of daily movements in exchange rates. Evans and Lyons focus on the US Dollar-German Mark and US Dollar-Japanese Yen exchange rates for May 1-Aug 31, 1996, and find that order flow accounts for 60 percent of the daily changes in the German exchange rates and 40 percent of the changes in the Yen. Brandt and Kavajecz (2004) focus on the US Treasury market, finding an effect of flows on yields that is large and strongest when liquidity is low. Finally, Ambrose, Lee, and Peek (2007) explore comovement in the REIT (Real Estate Investment Trust) market, looking at an event study created when REITs were added to the S&P general indices. They find that not only do the REITs included in the S&P indices commove more strongly with those indices after inclusion, the non-included REITs also commove more strongly with the indices after inclusion as well.

In all of this literature there is a concern that index-based comovement in returns reflect fundamentals, rather than common trading-induced price pressure. Our research is somewhat difficult: the inclusion in and especially deletions from the CDX indexes are driven by corporate events in direct way. Downgrades in particular induce deletion from the CDX investment grade index, and changing patterns of comovement include some fundamental component. But the CDS market is also a derivative market based on the underlying bonds, and hence we are able to use the changes in spreads on these underlying bonds as a control from firm fundamentals. We find that CDS spread betas increase more than bond spread betas after inclusion, and viceversa after deletion. This finding provides strong evidence for non-fundamental-based comovement in the CDS market.

3.3 Related literature on bond and CDS markets

This paper is related to the growing literature on bond and CDS markets. Collin-Dufresne, Goldstein, and Martin (2001) investigate the patterns of credit spread changes. They show that, using proxies that measure changes in default probabilities and changes in recovery rates, they are able to explain about 25 percent of observed credit spread changes. They find that the residuals from these explanatory regressions are highly cross-correlated, and appear to be driven by a single common factor. One potential explanation for this common factor would be market flows into and out of credit markets. The authors' approach is different from ours: they focus on bonds, where we focus on CDS markets.

Longstaff et al (2005) use the market for CDS to estimate the default and non-default components of corporate bond spreads. Their research uses the CDS spread to construct the true default probability of a corporate issuer, and apply that estimated default probability to corporate bonds to parse out the default and non-default related parts of bond spreads. They find that their measures of 'default probability' explains that bulk of bond spreads, but that a sizable part remains unexplained. Exploring the unexplained component of bond yields, they find that bond liquidity is an important determinant. Our paper is starting from an entirely different point – in showing patterns of CDS comovement around the inclusion and deletion of CDS issuers from the major indices, we are showing evidence of a liquidity-based component in the movements of these spreads.

3.4 Empirical design: inclusion in and deletion from the CDX indexes

CDS contracts are bilateral contracts used to transfer the risk of a 'credit event' between market participants. The 'protection seller' sells insurance to the 'protection buyer.' For single-name CDS contracts, the risk transferred is the risk of a credit event, typically a default, by a single issuer. This issuer can be a corporate or sovereign issuer, or an ABS. By transferring the risk of a credit event, credit default swaps accomplish a function that parallels the purchase of a physical bond; just as the purchaser of a physical bond holds the risk that the bond will default, the seller of protection under a credit default swap contract takes on an economically similar exposure.

The seller of credit protection is compensated by the payment of a credit spread, measured as some percentage of the notional value. This credit spread has always been regarded as a pure measure of the credit risk of the underlying reference entity, unpolluted by interest rate risk.

The CDS market has grown explosively over the past 10 years, with the notional single-name CDS exposure now exceeding the total notional value of the corporate bond market. As CDS contracts are traded in over-the-counter (OTC) markets rather than on exchanges, the market centers around a handful of major dealers. Pricing, although somewhat opaque, is available from sources such as Markit and CMA. The first indices of credit derivatives were created in 2001, and by 2004 the major index administrators (Trac-x and iBoxx) had merged to create the CDX indexes for North American credit and the iTraxx indexes for Europe. Markit Partners acquired both sets of indices in 2007, and is currently the administrator for all of the major credit derivative indexes.

There are a variety of different indexes covering different market subsegments. The North American market is covered by the CDX indexes: the Investment Grade (IG) index, the HVol subindex of the IG universe (HVol), the Crossover index, and the High Yield index and subindexes, and the sector-based indexes. There are also CDX Emerging market indexes. The iTraxx indexes, also owned by Markit, include European, Asian, and Australian markets. Additional credit indexes cover asset backed securities (the ABX, CMBX, and TABX), loans (the LCDX and LevX), sovereign debt (the SovX), and municipal securities (the MCDX).

Table 3.1 describes the current outstanding single-name and index credit derivatives contracts that were outstanding and registered with the Depository Trust Clearing Corporation (DTCC) as of May 2010. The DTCC registers the vast majority of all CDS contracts traded. The table shows the gross notional and net notional outstanding, as well as the total number of contracts. Many firms have offsetting positions in underlying instruments: the net notional provides a picture aggregating institutions net exposure. The CDX North America Investment Grade indexes, alongside the similar index for Europe, have the highest total outstanding gross and notional amounts, with outstanding amounts that are many times the next nearest contracts. Other index products are the most heavily traded individual instruments. Among single-name CDS contracts, the most heavily traded instruments are contracts referencing sovereign bonds. In particular, Italy, Turkey, Brazil, and Russia have large notional amounts.

There appears to be a discontinuous jump in the trading activity in CDS contracts that are included in the index versus contracts that are not included in the index. Causation works in both ways here: the dealer poll that drives index inclusion is based on selecting the more liquid and active CDS contracts for the index. At the same time, inclusion in the index drives trading related to index flows and products. Table 3.2 and table 3.3 show the magnitude of the activity discontinuity for names included in the index. Table 3.2 includes only the corporations among the top 1000 CDS reference entities in terms of trading activity, for a total of 442 firms. Trading activity is based on gross notional outstanding (columns 1-3), net notional outstanding (columns 4-6), and the number of contracts outstanding as of September 3, 2010.

There is a strong relationship between CDS trading activity and the amount of debt outstanding. Controlling for this relationship, though, inclusion in the CDX.IG index is associated with $9 Billion more gross outstanding in CDS contracts. Again, causation works both ways in this relationship, with inclusion in the index also being a reflection of underlying trading activity. Table 3.3 repeats the analysis of table 3.2, but fitting Tobit regressions using the entire sample of Compustat firms, with a truncation point set to the minimum value of each activity measure observed among the top 1000 issuers. The results are qualitatively similar, but the much larger coefficients on the CDX inclusion dummy variables reflect the truncated nature of the sample used in Table 3.2.

Table 3.4 shows the constituents for the most recent series (Series 14) of the CDX North American Investment Grade index. The constituents are chosen every 6 months by a poll of dealers, and as the name suggests are required to be investment-grade firms domiciled in North America. Table 3.5 shows the index additions and deletions for the recent rolls of the index. Deletions from the investment grade index commonly occur because of downgrades, but also follow mergers. In the case of Wells Fargo, a merger with Wachovia made Wells Fargo a CDX market maker, hence not eligible for inclusion in the index.

We use these periodic rolls of the CDX index to investigate patterns of comovement in the CDS market. Our hypothesis is that on inclusion in the index, the CDS spreads of an issuer will commove more with the average spreads in the index, due to the impact of correlated trading in index-based products and correlated hedging of index exposures. Specifically, both the beta of the spread on the index, as well as the R-squared, will go up.

3.5 Data

The main sample consists of CDS spreads of corporate issuers available from Datastream, which sources CDS data from CMA. CMA is a major provider of OTC market data, and along with Markit is the dominant provider of data on CDS spreads. We consider CDS spreads of issuers that were added or deleted from the CDX North America Investment Grade Index (CDX.NA.IG hereafter) between September 2004 and March 2009. The index inclusion and deletion dates for individual issuers are based on the sequence of constituents of the different series of the CDX.NA.IG index. The constituents of each of the CDX Index series are provided by Markit.com.

CDS contracts are written for a variety of different maturities, with 1,3,5,7, and 10 year contracts being the most common. Among these, the 5 year contracts are generally the most active and liquid and often viewed as the benchmark contracts for the issuer. We use the Datastream-reported spreads on the 5 year contracts in the analysis that follows. Because there are two main sources of data, we also show that the results are robust to the use of the CDS data provided by Markit. Relevant literature in CDS uses both sources of data. Although Markit has been widely considered as a more accurate source for CDS data, recent papers use CMA as the main source (see Bongaerts, Driessen, and De Jong, 2011, and Giglio, 2011). A recent study by Mayordomo, Peña, and Schwartz (2010) compares the major sources of corporate CDS prices and concludes that CMA database quotes lead the price discovery process in comparison with the quotes provided by other databases.

Data on the bonds matched to the CDS reference entities also come from Datastream, with the asset swap spread used as the primary measure of the bond spread. The asset swap spread reflects the equivalent spread over a floating-rate benchmark of a bond whose cash flows have been swapped from fixed to floating. This spread benchmark removes the direct impact of interest rate movements and is conceptually the closest match to the reported spread on a CDS contract, which also primarily reflects credit risk rather than interest rate effects. CDS are matched to the underlying bonds, with an algorithm used to select a liquid bond closest to the 5-year point. Data on the time series of the CDX.NA.IG comes from Bloomberg.

The total number of issuers that were included or deleted from the index ascends to 120. For an issuer to be included in our sample it has to be added to or deleted from the CDX.NA.IG between September 2004 and March 2009, and we also require a minimum of 80% of trading days per regression estimated. The final sample of issuers after the screening amounts to 95. There are 51 additions and 54 deletions that match our criteria. There are 10 issures that are both added to and deleted from the index in different rolls of the index.

Tables 3.6, 3.7, and 3.8 provide some descriptive statistics for the sample used in the paper. Table 3.6 aggregates the period between 2004 and 2010, while table 3.7 shows statistics for the pre-crisis period (up to July 2007), and table 3.8 shows the post-crisis period (after July 2007).

3.6 Results

To test our hypothesis, we run two regressions for each CDS issuer that has been included or excluded from the index, one the year before the event (the 255 trading days before the event), and another one the year after the inclusion (255 trading days after the event). For each issuer we regress the change in CDS spread on the change in the CDX spread:

$$\Delta CDS_{i,t} = \alpha_i + \beta_{ci} \Delta CDX_t + \varepsilon_{i,t}$$

We then compute the difference between the beta after the event and beta before the event, and label it $\Delta\beta_{ci}$, where the subindex c denotes CDS and i the issuer. The hypothesis predicts that the average change in beta, $\overline{\Delta\beta_c}$ should be significantly positive after an inclusion in the index, as well as the average change in the $\overline{\Delta R^2}$.

As mentioned in our identification strategy, we need to control for fundamentals, and we do so by computing the change in betas for the Asset Swap Spread (ASP) of the underlying bonds identified as the specific reference obligations of the CDS contract:

$$\Delta ASP_{i,t} = \alpha_i + \beta_{bi}\Delta CDX_t + \varepsilon_{i,t}$$

Before showing the results, it is important to understand the distribution of our data. CDS contracts only are widely available since 2004, this is why our sample spams only for 6 years. Table 3.6, 3.7, and 3.8 show the summary statistics for our full sample, the pre-crisis sample, and the crisis sample, respectively.

If we have a closer look at table 3.6, panel A (all observations), three aspects are worth noticing. First, there is a lot of variability in the CDS spreads during the whole period, with an average CDS spread of 284 and a median of 110. The sample is skewed positively. Second, we observe a very similar average and summary statistics for the bonds underlying, except at the very tail of the distribution. This confirms the fact that both assets are tied to the same issuer and should reflect the same credit risk. Third, we see that the median for changes in spread at the daily and weekly frequency is zero. As a consequence, in panel B we show the summary statistics for the observations where the change in daily CDS spreads is not zero. The number of such cases is not negligible, however it does not compromise our analysis, because the results are robust to this subsample of observations. For the full sample, as we can see comparing the column "Obs" for observations in the two panels, there is a 12% of observations for which there is no change in daily CDS spreads.

Tables 3.7 and 3.8 show the same statistics for the pre-crisis and crisis subsamples. A clear manifestation of the crisis was the high levels of CDS spreads for many corporate issuers. It is therefore important to show how the distribution of the main variables change for the different subsamples. In short, the mean and median of CDS spreads for the pre-crisis period were 129.87 and 84.10 respectively. The average CDS spread was more than tripled during the crisis period, to 482.22, and the median CDS spread was doubled to 163.20. The distribution became more skewed during the crisis period.

3.6.1 Additions

Table 3.9 shows the first set of results of our tests. In panel A we show the results using daily spread changes. Average betas of CDS spread changes are significantly higher after the addition than before the addition. For the full sample we see that the average change in beta for CDS amounts to 0.211 and is significant at the 1% level. The asterisks in the table reflect significance at the 1%, 5%, and 10% for one-sided tests, where we test whether the change in beta is bigger than zero. Because some of the additions take place in the same date, the standard errors are robust to cross-sectional correlation within addition dates. The average R-Squared also rises significantly after the addition by 0.040. However, this change in beta could be a consequence of the selection by the dealers poll. To account for changes in fundamentals of the issuer, we repeat the same exercise for the underlying bonds. If the change in betas for the changes in CDS spreads carry some information on the credit quality of the issuers, then it should be reflected as well in the changes in betas for the underlying bonds, and the difference in differences should not be significantly different from zero. We however find that the difference in differences of beta changes is a significant 0.307 with a standard error of 0.080. The same can be observed with the R-Squared, that has a difference in differences of 0.05, significant at the 5% level.

An important question raises when considering the sample period we use: is this effect being driven by the large increase of CDS spreads during the recent crisis? We find that the answer to that question is no. The effect that we document does not hinge in the great variability of CDS spreads of corporations during the crisis, rather in the increased attention and trading patterns of CDS index products. Our results confirm that this is the case. We then divide the sample in two subsamples, labeled "pre-crisis" (2004-2006) and "crisis" (2008-2010). We avoid using additions for which we need data both before the crisis and during the crisis to better disentangle the effect. Specifically, additions that occurred in March 2007 and September 2007 are not included in the pre-crisis nor in the crisis period, because the beta estimated before the addition will mainly contain data before the crisis whereas the post-event beta will use crisis period data.

Interestingly, the difference in difference results are stronger for the pre-crisis subsample than for the crisis subsample. The difference in differences in changes in bega for the pre-crisis period is 0.435 estimated accurately with a standard error of 0.134, whereas the difference in differences for the crisis period is 0.237 with a standard error of 0.145. The difference in difference is strong and significant in the pre-crisis sample because the change in betas for the underlying bonds was negative, while there is not a clear patter for the CDS change in beta. On the contrary, for the crisis sample, it is the beta in the CDS that is significantly positive and the underlying bond insignificant.

In panel B of the same table we show the results using weekly (Wednesday) spread changes, instead of daily, to mitigate the tradeoff between market microstructure effects when using high-frequency data and the statistical power of the tests. The change in betas for CDS spreads remains for the three sample periods, but the magnitude is bigger when using weekly data. The results are very robust to the use of weekly data, suggesting that the frequency with which we measure beta does not influence the results much.

These results point out at the clear existence of an excess-comovement triggered by the inclusion of a CDS into the CDX index that is not driven by fundamentals. The mechanisms underlying this comovement are discussed in the fourth subsection.

3.6.2 Robustness to sample of liquid observations

Although the companies that are included in the index tend to be very liquid, there are still companies for which there is no change in daily spread for more than one day. As explained above, there is a 12% of observations for which there is no change in the daily CDS spread. One could worry that the results might be driven by the lack of liquidity and the zero observations could affect this change in betas. To show that our results are not driven by this lack of variation in some instances, we repeat the analysis but using only observations for which there is a change different from zero in the daily CDS spread.

This results are shown in table 3.10. Results are by and large unchanged. Magnitudes are in line with thouse found in the benchmark specification. The difference in difference for the pre-crisis period is now 0.426 estimated accurately with a standard error of 0.196. The results are thus not driven by a lack of variation in CDS spreads, but rather remain strong and significant using a subsample of non-zero CDS spread changes.

3.6.3 Robustness to Markit database

It is important to test the robustness of the results with a different database, as Markit is the major vendor of CDS data. Markit has been widely considered as a more accurate source for CDS data, however recent papers use CMA as the main source (see Bongaerts, Driessen, and De Jong, 2011, and Giglio, 2011). A recent study by Mayordomo, Peña, and Schwartz (2010) compares the major sources of corporate CDS prices and concludes that CMA database quotes lead the price discovery process in comparison with the quotes provided by other databases.

In table 3.11 we show the results when using a different dataset for CDS spreads, Markit. Only 35 of the 38 benchmark additions could be matched with Markit database. All the results seem largely unchanged, with very small differences. Difference in differences for weekly returns are still very accurately estimated in the pre-crisis period, with a significance at the 1% level for both the full sample and the pre-crisis sample, confirming that the pre-crisis effect is dominant in magnitude and significance over the crisis sample. Table 3.12 we show the results only using observations for which there is a non-zero change in daily CDS spread, and the patterns are very similar to the ones in table 3.10.

3.6.4 Dimson betas

Previous research on comovement in the stock market attempts to dissentangle the sources of the observed change in comovement. According to Barberis, Shleifer, and Wurgler (2005), three are the possible sources of friction- or sentiment-based comovement, namely, category view, habitat view, and information diffusion. The category view, initially proposed by Barberis and Shleifer (2003), argues that investors tend to simplify portfolio decisions by allocating funds at the category level, instead of at the asset level. In the presence of noise traders with correlated sentiment that can affect prices, there appears an excess comovement into each category by moving funds from one to another group. Habitat view reflects the fact that many investors have a limited investment universe (a preferred habitat), due to transaction costs, or lack of information. This creates a common factor in the returns of these assets that is non-fundamental. Finally, the information difussion predicts that, due to market frictions, the information is incorporated quicker into the prices of some stocks than others.

The use of Dimson (1979) betas allows us to test whether the excess comovement is just a change in speed at which information is incorporated (due to market frictions), or else comes from a more sentiment-driven explanation such as category view or habitat view. We can do so by including leads and lags of the index in the daily analysis, to see if individual CDS react with "less" delay after being included in the index. We specifically run the following regression before and after each inclusion or deletion event:

$$\Delta CDS_{i,t} = \alpha_i + \sum_{s=-5}^{5} \beta_{ci}^{(s)} \Delta CDX_{t+s} + \varepsilon_{i,t}$$

and then we compute the difference between the sum of Dimson betas after the event and the sum of Dimson betas before the event. We then average them clustering for cross-sectional correlation. Similarly, to control for fundamentals, we estimate the same regression for the changes in asset swap spread:

$$\Delta ASP_{i,t} = \alpha_i + \sum_{s=-5}^{5} \beta_{bi}^{(s)} \Delta CDX_{t+s} + \varepsilon_{i,t}$$

This difference will give us then the change in comovement that would happen if there were no information difussion effects. In other words, if the effect disappears, then the excess comovement found in the previous section comes from the information difussion channel. If, however, there still remains a significant change in comovement, that would be evidence of an effect coming from the two other channels.

Empirical evidence on the importance of the information diffusion channel is mixed. Using this Dimson betas approach, Barberis et al. find that most of the excess-comovement associated with an S&P 500 index inclusion comes from an information diffusion explanation. However, a recent study by Green and Hwang (2009) shows that the excess-comovement that arises after a stock-split not only comes from information diffusion but from a pure category or habitat based explanation.

Table 3.15 shows that in the CDS market, information difussion is not driving our results. Results actually become even stronger than when using a single beta, as in the previous section. In Panel A we show the differences in betas after addition, where the betas are not single betas, but the sum of the 11 Dimson betas (current, plus 5 leads and 5 lags). For the full-sample, we observe that the change in Dimson beta for CDS is a significant 0.515 (compared to the 0.211 from a single beta, in table 3.9), and once controlled for the change in the associated betas from the bonds, it still remains a significant 3.58 (compared to the 0.307 from table 3.9). Panel B shows the composition of Dimson betas, and helps understand the results from Panel A. All the betas for CDS except two are positive, whereas five betas for the bonds are negative. The contemporaneous effect is very strong for the CDS and not for the bond. Table 3.16 shows that the results are by and large unchanged if we use the alternative Markit database.

These results strongly suggest that the category and preferred habitat channels play an important role in explaining the changes in comovement of CDS contracts added to the CDX index.

3.6.5 Deletions

In this subsection we comment on the results that come from deletions from the CDX index. Deletions from the index are in most cases a consequence of a downgrade in the underlying bond, or a merger of the company with another one already in the index. However, because we do test jointhly changes in betas for CDS spreads as well as the underlying bonds, these results are also relevant for our study.

Table 3.13 shows three main findings related to deletions using the full sample. First, changes in betas for CDS spreads are slightly negative, but not significantly different from zero. Second, there is a positive change in beta for the underlying bonds, especially using weekly spread changes. The intuition for this result is as follows. When the downgrade is announced, CDS spreads become more sensitive to changes in the CDX Index spread, and hence the beta before deletion is already high. With the downgrade, firms approach distress and their bonds begin to take on a larger share of the company's risk, so the underlying bonds beta also experience an increase. However, after deleltion, not-belonging to the index causes the comovement of the CDS spreads of the downgraded company to drop more than that of the underlying bonds, which were not linked to the CDX index. For weekly returns is especially clear. The change in beta for CDS spreads is -0.075 poorly estimated with a standard error of 0.256, whereas the change in beta for the underlying bonds is 0.440 with a standard deviation of 0.156. The diference in diferences is however not significant.

3.7 Conclusion

By exploring additions and deletions of corporate CDS into the CDX Index we provide evidence of an excess co-movement in CDS markets not driven by fundamental reasons. Many mutual funds and exchange traded funds are explicitly tied to these benchmark indexes. The flow of investors' money into and out of these funds induces correlation in trading activity across the index constituents.

To control for fundamentals we propose the novel approach of comparing changes in betas of CDS around inclusions with changes in betas of the underlying bonds. Because bonds and CDS contracts both offer investors economic exposure to an issuers' credit risk, their variation in a frictionless and unsegmented market should be parallel. We find that average changes in betas for CDS exceed significantly average changes in beta for the underlying bonds. We estimate Dimson betas, and find that the excess-comovement is not driven by an information diffusion channel, but induced by a category and preferred habitat channel.

We also show that deletions from the index see no statistically significant change in the mean beta of the CDS on the index, whereas changes in betas for the underlying bonds do. The betas are high prior to deletions because issuers being deleted from the CDX Investment Grade index are often being removed because they lose their investment grade status.

In net these results suggest that the markets for CDS and their underlying bonds are somewhat segmented, and that there is an excess co-movement among the CDS spreads that belong to the major CDX Index, the North American Investment Grade.

3.8 Tables

Table 3.1: Index and Single-Name CDS contracts

These are contracts registered with the Depository Trust & Clearing Corporation's Trade Information Warehouse (the 'DTCC Warehouse'), reported as of May 7, 2010. Gross notional and net notional amounts are in Billions of USD.

Indexes and Index tranches	Gross notional	Net notional	Contracts
CDX North American Investment Grade index	3,955	361	20,002
iTraxx Europe main index	3,362	424	11,033
CDX North American High Yield indexes	672	78	1,785
iTraxx Europe sector indexes	489	73	27
ITraxx Europe crossover index	390	36	547
CMBX indexes	194	35	28
iTraxx Europe HiVol index	182	37	113
iTraxx SovX indexes	181	13	1,328
Loan indexes	175	13	923
CDX.NA.IG.HVOL index	138	31	309
ABX and TABX indexes	137	28	60
CDX.EM index	108	18	461
iTraxx Asia ex-Japan Indexes	95	9	149
iTraxx Australia Index	94	8	623
iTraxx Japan index	65	10	53
CDX.NA.XO index	32	6	68
MCDX index	11	3	44
Total index	10,280	1,182	37,553

Single-name CDS contracts	Gross notional	Net notional	Contracts
Republic of Italy	216	24	5,537
Republic of Turkey	173	5	11,576
Federative Republic of Brazil	147	13	11,120
Russian Federation	115	4	8,383
United Mexican States	104	6	8,715
Kingdom of Spain	101	14	4,240
JPMorgan Chase & Co	84	5	9,239
General Electric Capital	83	11	7,690
Bank of America Corporation	82	6	9,191
Hellenic Republic (Greece)	75	8	3,645
Total single name	14,637	1,220	2,152,319

Table 3.2: Linear regressions of CDS trading activity, September 3, 2010

The results of linear regressions of CDS trading activity on firm fundamentals and CDS contract index status. CDS trading activity based on DTCC list of top 1000 CDS contracts, September 3, 2010. Firm fundamentals reflect most recent annual totals avaiable from Compustat. Sample includes only corporations among top 1000 global CDS contracts. Standard erros are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. variable	Gross value of contracts outstanding ($M)			Net value of contracts outstanding ($M)			Contracts outstanding		
Indep. variable									
CDX.IG index	9,115.316***		9,171.626***	546.194***		473.305***	1,221.678***		1,271.235***
	(1,228.368)		(834.148)	(102.706)		(61.778)	(137.295)		(111.609)
CDX.HY index	1,456.137		2,970.352***	-130.087		-3.568	419.590***		559.357***
	(1,371.104)		(917.778)	(114.641)		(67.971)	(153.249)		(122.798)
CDX.XO index	7,442.850***		8,495.265***	376.646**		471.332***	1,317.704***		1,411.149***
	(2,037.349)		(1,355.593)	(170.347)		(100.396)	(227.716)		(181.378)
Accounts payable		0.001	0.011*		-0.002***	-0.002***		0.001	0.003***
		(0.007)	(0.006)		(0.000)	(0.000)		(0.001)	(0.001)
Long-term debt		0.245***	0.227***		0.027***	0.026***		0.020***	0.017***
		(0.020)	(0.017)		(0.001)	(0.001)		(0.003)	(0.002)
Sales		-0.001	-0.010		0.001*	0.001		-0.001	-0.002
		(0.011)	(0.010)		(0.001)	(0.001)		(0.001)	(0.001)
Constant	8,837.763***	9,469.731***	6,008.521***	790.916***	650.865***	507.335***	1,538.593***	1,840.312***	1,304.033***
	(719.173)	(459.795)	(518.257)	(60.132)	(31.793)	(38.383)	(80.382)	(63.335)	(69.343)
Observations	442	442	442	442	442	442	442	442	442
R-squared	0.127	0.484	0.617	0.077	0.627	0.682	0.195	0.277	0.494

Table 3.3: Tobit regressions of CDS trading activity, September 3, 2010

The results of tobit regressions of CDS trading activity on firm fundamentals and CDS contract index status. CDS trading activity based on DTCC list of top 1000 CDS contracts, September 3, 2010. Firm fundamentals reflect most recent annual totals avaiable from Compustat. Sample includes all corporations in Compustat with valid accounts payable, debt, and sales variables. Truncation point for tobit model set to minimum value among DTCC top 1000 list. Standard erros are in parentheses. * significant at 10%; ** significant at 5%; *** significant at 1%.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Dep. variable	Gross value of contracts outstanding ($M)			Net value of contracts outstanding ($M)			Contracts outstanding		
Indep. variable									
CDX.IG index	47,755.5***		37,409.3***	4,005.8***		2,993.1***	6,729.3***		5,566.8***
	(1,988.3)		(1,575.1)	(176.6)		(135.8)	(277.4)		(235.3)
CDX.HY index	36,205.9***		30,355.2***	2,981.2***		2,398.6***	5,372.7***		4,793.2***
	(2,155.1)		(1,746.2)	(191.2)		(150.7)	(299.4)		(260.0)
CDX.XO index	26,075.6***		23,154.3***	2,044.9***		1,752.9***	3,973.5***		3,688.7***
	(3,042.3)		(2,486.8)	(269.4)		(214.3)	(419.1)		(366.5)
Accounts payable		-0.027***	0.005		-0.003***	-0.000		-0.005***	0.001
		(0.010)	(0.005)		(0.001)	(0.000)		(0.002)	(0.001)
Long-term debt		0.251***	0.150***		0.021***	0.014***		0.034***	0.017***
		(0.028)	(0.016)		(0.002)	(0.001)		(0.005)	(0.002)
Sales		0.293***	0.115***		0.023***	0.011***		0.047***	0.017***
		(0.022)	(0.013)		(0.002)	(0.001)		(0.004)	(0.002)
Constant	-29,802.4***	-37,879.0***	-25,242.3***	-2,668.7***	-2,924.8***	-2,206.0***	-3,969.0***	-6,006.0***	-3,579.9***
	(1,333.3)	(1,763.6)	(1,103.9)	(118.9)	(134.2)	(95.8)	(189.2)	(291.2)	(168.9)
Observations	9685	9685	9685	9685	9685	9685	9685	9685	9685

Table 3.4: Markit CDX.NA.IG index constituents, Series 14

Table shows issuer, average credit rating of bonds issued by entity, and industry classification.

ACE Ltd / A / Fin	Duke Energy / A / Ut	Ryder Sys Inc / A / Ind
Aetna Inc. / A / Fin	E I du Pont / A / Mats	Safeway Inc / BBB / Cons Stable
Alcoa Inc. / BBB / Mats	Eastman Chem Co / BBB / Mats	Sara Lee Corp / BBB / Cons Stable
Altria Gp Inc / BBB / Cons Stable	ERP Oper Ltd Pship / A / Fin	Sempra Engy / A / Ut
Amern Elec Pwr Co Inc / BBB / Ut	FirstEnergy Corp / BBB / Ut	Simon Ppty Gp L P / A / Fin
Amern Express Co / A / Fin	Fortune Brds / BBB / Cons Stable	SLM Corp / BBB / Fin
Amern Intl Gp Inc / BBB / Fin	Freeport McMoran / BBB / Mats	Southwest / BBB / Cons Cyc
Amgen Inc. / A / Cons Stable	G A T X Corp / BBB / Ind	Staples Inc / BBB / Cons Cyc
Anadarko / BBB / Energy	Gen Elec Cap Corp / AA / Fin	Target Corp / A / Cons Cyc
Arrow Electrs Inc / BBB / Ind	Gen Mls Inc / BBB / Cons Stable	Allstate Corp / BBB / Fin
AT&T Inc / A / Comm+Tech	Goodrich Corp / BBB / Ind	Black&Decker Corp / A / Ind
AT&T Mobility / A / Comm+Tech	Halliburton Co / A / Energy	Chubb Corp / A / Fin
Autozone Inc / BBB / Cons Cyc	Hewlett Pckd / A / Comm+Tech	Dow Chem Co / BBB / Mats
Avnet, Inc. / BBB / Ind	Honeywell Intl Inc / A / Ind	Hartford Finl / BBB / Fin
Barrick Gold Corp / BBB / Mats	Ingersoll Rand Co / A / Ind	Home Depot Inc / BBB / Cons Cyc
Baxter Intl Inc / A / Cons Stable	IBM Corp / A / Comm+Tech	The Kroger Co. / BBB / Cons Stable
Boeing Cap Corp / A / Fin	Intl Paper Co / BBB / Mats	Sherwin Williams Co / A / Cons Cyc
Boston Pptys / BBB / Not given	Johnson Ctls Inc / BBB / Ind	TJX Cos Inc / A / Cons Cyc
Bristol Myers / A / Cons Stable	Kinder Morgan / BBB / Energy	Walt Disney Co / A / Cons Cyc
Burlington Nthn / BBB / Ind	Kohls Corp / BBB / Cons Cyc	TIME WARNER C / BBB / Not given
Campbell Soup / A / Cons Stable	Kraft / BBB / Cons Stable	Time Warner Inc / BBB / Comm+Tech
Cdn Nat Res Ltd / BBB / Energy	Lockheed Martin Corp / A / Ind	Toll Bros Inc / BBB / Cons Cyc
Cap One / A / Fin	Loews Corp / A / Cons Stable	Transocean Inc / BBB / Energy
Cardinal Hlth/ BBB / Cons Cyc	Lowes Cos Inc / A / Cons Cyc	Un Pac Corp / BBB / Ind
Carnival Corp / A / Ind	M D C Hldgs Inc / BBB / Cons Cyc	Utd Parcel Svc Inc / AA / Ind
Caterpillar Inc / A / Cons Cyc	Marriott Intl Inc / BBB / Cons Cyc	UnitedHealth Gp Inc / A / Fin
CBS Corp / BBB / Cons Cyc	Marsh&Mclenn / BBB / Fin	Unvl Health / BBB / Cons Stable
CenturyTel / BBB / Comm+Tech	McDonalds Corp / A / Cons Cyc	Valero Energy Corp / BBB / Energy
Cigna Corp / BBB / Fin	McKesson Corp / BBB / Cons Cyc	Verizon / A / Comm+Tech
Cisco Sys Inc / A / Comm+Tech	MetLife Inc / A / Fin	Viacom / BBB / Not given
Comcast / BBB / Comm+Tech	Motorola Inc / BBB / Ind	Vornado Rlty LP / BBB / Fin
Comp Sci / BBB / Comm+Tech	NRUC / A / Ut	Wal Mart / AA / Cons Cyc
ConAgra / BBB / Cons Cyc	Newell Rubbmd. / BBB / Ind	Whirlpool Corp / BBB / Cons Cyc
ConocoPhillips / A / Energy	News Am / BBB / Comm+Tech	Xerox Corp / BBB / Cons Cyc
Const Engy Gp / BBB / Ut	Nordstrom Inc / A / Cons Cyc	XL Cap Ltd / BBB / Fin
Cox / BBB / Comm+Tech	Norfolk Sthn Corp / BBB / Ind	XTO Engy Inc / BBB / Energy
CSX Corp / BBB / Ind	Northrop Grumm / BBB / Ind	YUM Brands Inc / BBB / Cons Cyc
CVS / BBB / Cons Cyc	Omnicom Gp Inc / A / Comm+Tech	
Darden Rest / BBB / Cons Cyc	Pfizer Inc / AA / Cons Stable	
Deere&Co / A / Cons Cyc	Progress Engy Inc / BBB / Ut	
Dell Inc / A / Comm+Tech	Quest Diagnostics Inc / BBB / Ind	
Devon Engy Corp / BBB / Energy	R R Donnelley / BBB / Comm+Tech	
DIRECTV / BBB / Comm+Tech	Raytheon Co / A / Ind	
Dominion Res Inc / BBB / Ut	Reynolds A Inc / BBB / Cons Stable	

Table 3.5: Additions to and deletions from Markit CDX, Series 12-14

This table shows the additions to and deletions from Markit CDX for series 12 to 14, as well as the reason for removal.

Series	Announc date	Removed	Reason for removal	Added
14	Mar-10	International Lease Finance Corporation	Downgrade	Freeport-McMoRan Copper and Gold Inc.
		Wells Fargo and Company	Merger with Wachovia*	SLM Corporation
13	Sep-09	CIT Group Inc.	Downgrade/Bankruptcy	DIRECTV HOLDINGS LLC
		J. C. Penney Company, Inc.	Downgrade	GATX CORPORATION
		Macy's, Inc.	Downgrade	JOHNSON CONTROLS, INC.
		Masco Corporation		KINDER MORGAN ENERGY PARTNERS, L.P.
		Textron Financial Corporation	Downgrade	REYNOLDS AMERICAN INC.
		Weyerhaeuser Company		UnitedHealth Group Incorporated
12	Mar-09	Gannett	Downgrade	Avnet
		iStar Financial	Downgrade	Boston Properties
		Limited Brands	Downgrade	Canadian Natural Resources
		The New York Times	Downgrade	Cisco Systems
		MBIA Insurance	Downgrade	Dell
		Mohawk Industries	Downgrade	Lowe's
		Starwood Hotels&Resorts	Downgrade	Pfizer
		Wyeth	Merger with Pfizer	TJX Co.
		Embarq	Merger with CenturyTel	Vornado Realty

Table 3.6: Descriptive statistics on CDS and bond returns, full sample

This table shows the descriptive statistics on CDS and bond returns for the full period of our sample: January 2004 to March 2010. Panel A displays the summary statistics of CDS and bond returns for all observations, whereas Panel B shows only summary statistics of those observations for which the change in daily CDS spread is not zero.

PANEL A: ALL OBSERVATIONS (Full sample period: January 2004 to March 2010)

CDS	Frequency	Obs	Mean	SD	Min	1th ptile	5th	25th	50th	75th	95th	99th	Max
Spread	Daily	56151	283.99	685.49	9.50	20.00	31.55	62.50	110.00	234.70	858.30	4019.03	12660.70
Change in spread	Daily	56087	0.54	76.10	-5678.20	-56.95	-13.79	-1.85	0.00	2.00	15.92	67.50	6527.20
Change in spread	Weekly	56043	2.81	139.44	-5145.80	-165.20	-41.40	-5.20	0.00	6.80	49.20	207.50	6907.70

Bonds	Frequency	Obs	Mean	SD	Min	1th ptile	5th ptile	25th	50th	75th	95th	99th	Max
Spread	Daily	46845	291.13	380.80	-54.40	16.60	47.00	94.80	165.00	344.90	901.80	2141.20	4062.20
Change in spread	Daily	46819	0.45	21.12	-646.80	-47.00	-14.10	-3.30	0.10	3.40	15.70	53.70	968.20
Change in spread	Weekly	46715	2.26	47.87	-670.90	-123.70	-39.90	-6.40	0.20	7.70	48.80	167.80	953.60

PANEL A: NON-ZERO SPREAD CHANGES (Full sample period: January 2004 to March 2010)

CDS	Frequency	Obs	Mean	SD	Min	1th ptile	5th	25th	50th	75th	95th	99th	Max
Spread	Daily	49517	270.22	588.27	9.50	20.10	35.30	66.70	115.70	246.14	835.00	3434.60	12660.70
Change in spread	Daily	49453	0.62	81.05	-5678.20	-64.20	-15.45	-2.30	-0.01	2.60	17.91	73.80	6527.20
Change in spread	Weekly	49425	3.15	144.60	-5145.80	-179.40	-44.88	-6.00	0.20	8.00	53.55	227.50	6907.70

Bonds	Frequency	Obs	Mean	SD	Min	1th ptile	5th ptile	25th	50th	75th	95th	99th	Max
Spread	Daily	41751	298.48	389.35	-51.50	18.10	48.10	97.90	171.80	352.60	911.80	2203.90	4062.20
Change in spread	Daily	41726	0.50	21.62	-646.80	-48.60	-14.50	-3.40	0.10	3.50	16.40	55.90	968.20
Change in spread	Weekly	41637	2.51	48.86	-670.90	-124.60	-40.20	-6.60	0.30	8.00	50.70	171.20	953.60

Table 3.7: Descriptive statistics on CDS and bond returns, pre-crisis period

This table shows the descriptive statistics on CDS and bond returns for the pre-crisis period: January 2004 to July 2007. Panel A displays the summary statistics of CDS and bond returns for all observations, whereas Panel B shows only summary statistics of those observations for which the change in daily CDS spread is not zero.

PANEL A: ALL OBSERVATIONS (Pre-crisis sample period: January 2004 to July 2007)

CDS	Frequency	N	Mean	SD	Min	1th ptile	5th	25th	50th	75th	95th	99th	Max
Spread	Daily	27177	129.87	176.53	9.50	18.60	25.90	53.00	84.10	134.50	375.70	647.50	2607.40
Change in spread	Daily	27147	0.27	12.91	-731.80	-20.70	-6.80	-1.30	0.00	1.20	8.20	27.50	741.00
Change in spread	Weekly	27103	1.34	27.47	-704.70	-55.00	-17.50	-3.50	0.00	3.80	24.00	75.30	959.40

Bonds	Frequency	N	Mean	SD	Min	1th ptile	5th ptile	25th	50th	75th	95th	99th	Max
Spread	Daily	22228	132.32	93.83	-54.40	5.00	35.80	72.50	105.70	154.55	344.30	439.50	689.90
Change in spread	Daily	22206	0.14	7.89	-143.80	-19.70	-8.40	-2.60	0.00	2.40	9.00	23.60	199.80
Change in spread	Weekly	22118	0.70	16.91	-221.70	-43.10	-17.90	-4.60	-0.20	4.40	21.50	62.00	258.10

PANEL A: NON-ZERO SPREAD CHANGES (Pre-crisis sample period: January 2004 to July 2007)

CDS	Frequency	N	Mean	SD	Min	1th ptile	5th	25th	50th	75th	95th	99th	Max
Spread	Daily	23359	127.18	135.34	9.50	18.90	27.30	55.70	87.50	138.50	380.80	578.30	2607.40
Change in spread	Daily	23329	0.31	13.93	-731.80	-23.00	-7.50	-1.70	-0.20	1.70	9.40	30.00	741.00
Change in spread	Weekly	23296	1.74	28.46	-704.70	-56.10	-18.30	-4.00	0.00	4.60	26.50	82.30	959.40

Bonds	Frequency	N	Mean	SD	Min	1th ptile	5th ptile	25th	50th	75th	95th	99th	Max
Spread	Daily	19019	135.23	96.17	-51.50	6.70	36.20	72.80	107.10	158.40	349.40	442.80	689.90
Change in spread	Daily	18998	0.16	8.14	-143.80	-20.30	-8.60	-2.70	0.00	2.50	9.40	24.50	199.80
Change in spread	Weekly	18922	0.89	17.41	-221.70	-43.20	-18.20	-4.60	-0.10	4.50	23.00	65.80	258.10

Table 3.8: Descriptive statistics on CDS and bond returns, crisis period

This table shows the descriptive statistics on CDS and bond returns for the crisis period: August 2007 to March 2010. Panel A displays the summary statistics of CDS and bond returns for all observations, whereas Panel B shows only summary statistics of those observations for which the change in daily CDS spread is not zero.

PANEL A: ALL OBSERVATIONS (Crisis sample period: August 2007 to March 2010)

CDS	Frequency	N	Mean	SD	Min	1th ptile	5th	25th	50th	75th	95th	99th	Max
Spread	Daily	29015	428.22	914.98	13.50	26.80	37.50	80.00	163.20	361.20	1828.70	6055.40	12660.70
Change in spread	Daily	28981	0.79	105.14	-5678.20	-103.10	-21.70	-2.90	0.00	3.34	24.17	117.70	6527.20
Change in spread	Weekly	28981	4.22	192.08	-5145.80	-339.50	-65.00	-8.80	0.03	11.24	75.00	377.30	6907.70

Bonds	Frequency	N	Mean	SD	Min	1th ptile	5th ptile	25th	50th	75th	95th	99th	Max
Spread	Daily	24655	434.13	473.76	-45.10	43.70	76.80	164.00	273.10	527.50	1282.20	2621.10	4062.20
Change in spread	Daily	24651	0.72	28.12	-646.80	-71.70	-19.80	-4.50	0.20	5.00	22.70	79.90	968.20
Change in spread	Weekly	24635	3.70	63.93	-670.90	-179.60	-63.30	-9.70	1.10	12.90	80.50	231.70	953.60

PANEL A: NON-ZERO SPREAD CHANGES (Crisis sample period: August 2007 to March 2010)

CDS	Frequency	N	Mean	SD	Min	1th ptile	5th	25th	50th	75th	95th	99th	Max
Spread	Daily	26198	397.66	776.74	13.50	26.80	41.94	87.95	170.17	365.20	1553.60	4162.59	12660.70
Change in spread	Daily	26164	0.88	110.65	-5678.20	-114.66	-23.68	-3.50	0.20	4.10	26.70	126.90	6527.20
Change in spread	Weekly	26169	4.44	196.91	-5145.80	-368.92	-68.70	-10.00	0.85	12.70	82.50	404.00	6907.70

Bonds	Frequency	N	Mean	SD	Min	1th ptile	5th ptile	25th	50th	75th	95th	99th	Max
Spread	Daily	22769	434.64	479.08	-45.10	44.30	79.30	165.70	272.00	523.80	1291.50	2649.00	4062.20
Change in spread	Daily	22765	0.78	28.31	-646.80	-72.20	-20.10	-4.50	0.30	5.10	23.10	81.40	968.20
Change in spread	Weekly	22752	3.89	64.14	-670.90	-178.30	-62.70	-9.70	1.20	13.00	80.50	231.90	953.60

Table 3.9: Changes in betas and R-Squares in CDS after addition to CDX

This table shows the average changes in estimated betas for changes in CDS spreads before and after the inclusion in the CDX.NA.IG Index. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Panel A reports results from the regressions using daily data, whereas Panel B shows results using weekly (Wednesday) data. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DAILY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	0.211***	0.040*	-.096	-.010	0.307***	0.050**
2004-2010		(0.081)	(0.027)	(0.074)	(0.005)	(0.080)	(0.027)
Pre-crisis	11	0.082	-.002	-.353	-.008	0.435***	0.006
2004-2006		(0.144)	(0.062)	(0.071)	(0.001)	(0.134)	(0.060)
Crisis	21	0.180**	0.049**	-.057	-.015	0.237*	0.064**
2008-2010		(0.094)	(0.028)	(0.071)	(0.010)	(0.145)	(0.029)

PANEL B: WEEKLY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	0.320***	0.040	0.054	-.009	0.266**	0.049
2004-2010		(0.133)	(0.042)	(0.161)	(0.009)	(0.158)	(0.043)
Pre-crisis	11	0.074	-.066	-.498	-.031	0.572***	-.035
2004-2006		(0.280)	(0.052)	(0.247)	(0.010)	(0.076)	(0.059)
Crisis	21	0.253**	0.049	0.175	0.004	0.078	0.045
2008-2010		(0.139)	(0.039)	(0.173)	(0.010)	(0.250)	(0.048)

Table 3.10: Results for additions, only non-zero daily spread changes

This table shows the average changes in estimated betas for changes in CDS spreads before and after the inclusion in the CDX.NA.IG Index, using only the observations for which the daily change in CDS spread is different from zero. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Panel A reports results from the regressions using daily data, whereas Panel B shows results using weekly (Wednesday) data. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DAILY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	0.168**	0.039*	-.121	-.011	0.290***	0.050**
2004-2010		(0.102)	(0.027)	(0.089)	(0.006)	(0.091)	(0.028)
Pre-crisis	11	-.031	0.003	-.457	-.006	0.426**	0.009
2004-2006		(0.192)	(0.064)	(0.068)	(0.001)	(0.218)	(0.065)
Crisis	21	0.167*	0.046*	-.056	-.017	0.224*	0.063**
2008-2010		(0.106)	(0.030)	(0.067)	(0.010)	(0.157)	(0.031)

PANEL B: WEEKLY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	0.279**	0.039	0.050	-.004	0.229*	0.043
2004-2010		(0.139)	(0.038)	(0.167)	(0.010)	(0.148)	(0.045)
Pre-crisis	11	-.017	-.046	-.541	-.019	0.524***	-.028
2004-2006		(0.227)	(0.050)	(0.228)	(0.020)	(0.116)	(0.068)
Crisis	21	0.211**	0.032	0.181	0.009	0.031	0.023
2008-2010		(0.126)	(0.041)	(0.170)	(0.014)	(0.226)	(0.054)

Table 3.11: Results for additions (Markit)

This table shows the average changes in estimated betas for changes in CDS spreads before and after the inclusion in the CDX.NA.IG Index, using a different source of data for CDS: Markit. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Panel A reports results from the regressions using daily data, whereas Panel B shows results using weekly (Wednesday) data. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DAILY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	35	0.204**	0.027	-.062	-.010	0.266***	0.037
2004-2010		(0.090)	(0.034)	(0.066)	(0.006)	(0.099)	(0.035)
Pre-crisis	8	0.073	0.003	-.302	-.004	0.376**	0.008
2004-2006		(0.175)	(0.057)	(0.126)	(0.001)	(0.178)	(0.058)
Crisis	21	0.170*	-.006	-.057	-.015	0.227*	0.009
2008-2010		(0.132)	(0.038)	(0.071)	(0.010)	(0.175)	(0.047)

PANEL B: WEEKLY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	35	0.305**	0.055	0.116	-.005	0.189	0.060
2004-2010		(0.152)	(0.049)	(0.153)	(0.009)	(0.177)	(0.054)
Pre-crisis	8	0.050	-.023	-.429	-.020	0.479***	-.003
2004-2006		(0.366)	(0.030)	(0.383)	(0.020)	(0.171)	(0.043)
Crisis	21	0.228	0.010	0.175	0.004	0.054	0.006
2008-2010		(0.188)	(0.058)	(0.173)	(0.010)	(0.290)	(0.068)

Table 3.12:

Results for additions, only non-zero daily spread changes (Markit)

This table shows the average changes in estimated betas for changes in CDS spreads before and after the inclusion in the CDX.NA.IG Index, using a different source of data for CDS: Markit. We only use here the observations for which the daily change in CDS spread is different from zero. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Panel A reports results from the regressions using daily data, whereas Panel B shows results using weekly (Wednesday) data. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DAILY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	35	0.227***	0.033	-.093	-.010	0.320***	0.044*
2004-2010		(0.087)	(0.032)	(0.080)	(0.006)	(0.108)	(0.033)
Pre-crisis	8	0.162	0.026	-.459	-.002	0.621***	0.028
2004-2006		(0.177)	(0.051)	(0.092)	(0.005)	(0.148)	(0.055)
Crisis	21	0.176*	-.003	-.056	-.017	0.232*	0.014
2008-2010		(0.131)	(0.036)	(0.067)	(0.010)	(0.171)	(0.045)

PANEL B: WEEKLY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	35	0.305**	0.055	0.117	-.005	0.188	0.060
2004-2010		(0.152)	(0.049)	(0.153)	(0.009)	(0.177)	(0.054)
Pre-crisis	8	0.049	-.023	-.430	-.020	0.478***	-.003
2004-2006		(0.366)	(0.030)	(0.383)	(0.021)	(0.171)	(0.044)
Crisis	21	0.228	0.010	0.174	0.004	0.055	0.006
2008-2010		(0.188)	(0.058)	(0.172)	(0.010)	(0.291)	(0.068)

Table 3.13: Changes in betas and R-Squares after deletions from the CDX

This table shows the average changes in estimated betas for changes in CDS spreads before and after the deletion from the CDX.NA.IG Index. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Panel A reports results from the regressions using daily data, whereas Panel B shows results using weekly (Wednesday) data. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DAILY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	-.076	0.003	0.177	0.009	-.252	-.007
2004-2010		(0.187)	(0.030)	(0.148)	(0.015)	(0.248)	(0.026)
Pre-crisis	9	0.123	0.009	-.250	-.019	0.373***	0.028
2004-2006		(0.288)	(0.076)	(0.413)	(0.035)	(0.125)	(0.043)
Crisis	18	-.204	-.026	0.118*	-.009	-.322	-.019
2008-2010		(0.391)	(0.057)	(0.089)	(0.010)	(0.468)	(0.059)

PANEL B: WEEKLY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	-.093	-.044	0.424***	0.026***	-.517	-.070
2004-2010		(0.284)	(0.081)	(0.129)	(0.011)	(0.374)	(0.078)
Pre-crisis	9	-.239	-.008	0.509**	0.008	-.747	-.016
2004-2006		(0.231)	(0.093)	(0.278)	(0.029)	(0.495)	(0.068)
Crisis	18	-.173	-.124	0.406*	0.023**	-.579	-.146
2008-2010		(0.664)	(0.117)	(0.289)	(0.012)	(0.864)	(0.131)

Table 3.14: Results for deletions, only non-zero daily spread changes

This table shows the average changes in estimated betas for changes in CDS spreads before and after the deletion from the CDX.NA.IG Index, using only the observations for which the daily change in CDS spread is different from zero. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Panel A reports results from the regressions using daily data, whereas Panel B shows results using weekly (Wednesday) data. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DAILY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	-.114	-.002	0.198*	0.012	-.312	-.014
2004-2010		(0.190)	(0.031)	(0.138)	(0.015)	(0.259)	(0.028)
Pre-crisis	9	0.118	0.007	-.242	-.021	0.359***	0.027
2004-2006		(0.280)	(0.079)	(0.339)	(0.035)	(0.060)	(0.045)
Crisis	18	-.240	-.031	0.169*	-.001	-.409	-.029
2008-2010		(0.376)	(0.057)	(0.104)	(0.014)	(0.463)	(0.063)

PANEL B: WEEKLY SPREAD CHANGES

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	-.075	-.047	0.440***	0.023**	-.516	-.070
2004-2010		(0.256)	(0.077)	(0.156)	(0.013)	(0.359)	(0.074)
Pre-crisis	9	-.144	-.000	0.530**	-.004	-.674	0.004
2004-2006		(0.223)	(0.083)	(0.228)	(0.030)	(0.410)	(0.058)
Crisis	18	-.139	-.127	0.461	0.020*	-.599	-.147
2008-2010		(0.625)	(0.112)	(0.370)	(0.016)	(0.861)	(0.126)

Table 3.15: Changes in CDS Dimson betas after addition to the CDX

In Panel A we show the average changes in the sum of up five leads and lags of estimated betas (Dimson betas) before and after the deletion from the CDX.NA.IG Index. In Panel B we show each of the components of the Dimson betas. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DIMSON BETA

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	38	0.515**	0.036	0.117	-.012	0.398***	0.048**
2004-2010		(0.225)	(0.039)	(0.229)	(0.016)	(0.170)	(0.027)
Pre-crisis	11	0.174	-.058	-.577	-.050	0.752***	-.008
2004-2006		(0.469)	(0.084)	(0.158)	(0.028)	(0.312)	(0.064)
Crisis	21	0.569*	0.067***	0.397	0.004	0.172**	0.062***
2008-2010		(0.362)	(0.014)	(0.320)	(0.013)	(0.104)	(0.003)

PANEL B: COMPONENTS OF DIMSON BETA

Full sample	$t-5$	0.064	0.114***	-.050
2004-2010		(0.067)	(0.041)	(0.063)
	$t-4$	0.006	0.083	-.078
		(0.032)	(0.077)	(0.061)
	$t-3$	-.051	-.133	0.082
		(0.105)	(0.138)	(0.077)
	$t-2$	0.051	0.135***	-.084
		(0.068)	(0.054)	(0.055)
	$t-1$	0.041	-.034	0.075
		(0.075)	(0.061)	(0.070)
	t	0.241***	-.078	0.320***
		(0.084)	(0.061)	(0.086)
	$t+1$	0.030	0.121***	-.091
		(0.067)	(0.051)	(0.063)
	$t+2$	-.048	-.050	0.002
		(0.021)	(0.063)	(0.054)
	$t+3$	0.017	0.117*	-.100
		(0.046)	(0.090)	(0.110)
	$t+4$	0.126**	0.018	0.108***
		(0.067)	(0.070)	(0.040)
	$t+5$	0.038*	-.176	0.214*
		(0.026)	(0.137)	(0.144)

Table 3.16: Changes in CDS Dimson betas (Markit)

In Panel A we show the average changes in the sum of up five leads and lags of estimated betas (Dimson betas) before and after the deletion from the CDX.NA.IG Index, for the Markit database. In Panel B we show each of the components of the Dimson betas. Reported coefficients show changes in betas and changes in R-Squares from 1 year estimation windows. Standard erros (in parenthesis) are robust to cross-setional correlation within cluster of additions. * significant at 10%; ** significant at 5%; *** significant at 1%, for one-sided tests, where the test is whether the coefficient is greater than zero.

PANEL A: DIMSON BETA

		CDS		Underlying Bond		Difference	
	N	$\Delta\beta_c$	ΔR_c	$\Delta\beta_b$	ΔR_b	$\Delta\Delta\beta$	$\Delta\Delta R$
Full sample	35	0.536***	0.020	0.146	-.003	0.390**	0.023
2004-2010		(0.210)	(0.031)	(0.233)	(0.011)	(0.171)	(0.032)
Pre-crisis	8	0.258	-.044	-.713	-.025	0.971***	-.020
2004-2006		(0.535)	(0.044)	(0.160)	(0.026)	(0.414)	(0.039)
Crisis	21	0.537**	-.002	0.397	0.004	0.140***	-.006
2008-2010		(0.317)	(0.013)	(0.320)	(0.013)	(0.042)	(0.014)

PANEL B: COMPONENTS OF DIMSON BETA

Full sample	$t-5$	0.077*	0.116***	-.039
2004-2010		(0.054)	(0.046)	(0.057)
	$t-4$	0.010	0.049	-.039
		(0.025)	(0.076)	(0.078)
	$t-3$	-.004	-.067	0.063
		(0.036)	(0.103)	(0.090)
	$t-2$	-.015	0.117**	-.131
		(0.033)	(0.051)	(0.051)
	$t-1$	0.102**	-.011	0.112**
		(0.050)	(0.045)	(0.060)
	t	0.218***	-.061	0.278***
		(0.087)	(0.062)	(0.093)
	$t+1$	0.043	0.137***	-.093
		(0.057)	(0.057)	(0.058)
	$t+2$	-.037	-.022	-.014
		(0.027)	(0.059)	(0.056)
	$t+3$	0.032	0.057	-.026
		(0.041)	(0.054)	(0.081)
	$t+4$	0.063*	-.036	0.098**
		(0.042)	(0.050)	(0.048)
	$t+5$	0.047**	-.133	0.180*
		(0.024)	(0.109)	(0.113)

References

Agarwal, A. and N.Y. Nail, 2004, Risk and Portfolio Decisions involving Hedge Funds, *Review of Financial Studies* 17, 63-98.

Ambrose, Brent, Dong Wook Lee, and Joe Peek, 2007, Comovement After Joining an Index: Spillovers of Nonfundamental Effects, *Real Estate Economics*, 35:1, pp. 57-90.

Antón, Miguel and Christopher Polk, 2010, Connected Stocks, London School of Economics working paper, FMG Discussion Paper 651.

Asness, Clifford S., Robert J. Krail, and John M. Liew, 2001, Do Hedge Funds Hedge?, *The Journal of Portfolio Management*, 28, 6-19.

Barberis, Nicholas, and Andrei Shleifer, 2003, Style Investing, *Journal of Financial Economics*, 68, pp. 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.

Blanco, Roberto, Simon Brennan, and Ian Marsh, 2005, An Empirical Analysis of the Dynamic Relation between Investment-Grade Bonds and Credit Default Swaps, *Journal of Finance* 60:5, pp. 2255-2281.

Bondarenko, Oleg, 2004, Market Price of Variance Risk and Performance of Hedge Funds, University of Illinois Chicago working paper.

Bongaerts, D.G.J., J.J.A.G. Driessen, and F.C.J.M. De Jong, 2011, Derivative Pricing with Liquidity Risk: Theory and Evidence from the Credit Default Swap Market, *Journal of Finance*, 66(1), 203-240.

Boyer, Brian, Tomomi Kumagai, and Kathy Yuan, 2006, How Do Crises Spread? Evidence from Accessible and Inaccessible Stock Indices, *Journal of Finance* 61:2, pp. 957-1002.

Brandt, Michael, and Kenneth Kavajecz, 2004, Price Disovery in the US Treasury Market: The Impact of Order Flow and Liquidity on the Yield Curve, *Journal of Finance*, 59, pp. 2623-2654.

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, Markus 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., Stefano Giglio, and Christopher Polk, 2010, Hard Times, London School of Economics working paper.

Campbell, John Y., and J. Mei., 1993, Where do betas come from? Asset price dynamics and the sources of systematic risk, *Review of Financial Studies* 6, 567-592.

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.

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

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

Chen, Nai-Fu, Richard Roll, and Steven A. Ross, 1986, Economic forces and the stock market, *Journal of Business* 59, 383-403.

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, 2010, The Prices is (almost) Right, *Journal of Finance* 64, 2739-2782.

Collin-Dufresne, P., Robert S. Goldstein, and Spencer J. Martin, 2001, The Determinants of Credit Spreads, *Journal of Finance*, 56, No. 6.

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*, forthcoming.

Da, Zhi, and Mitchell Warachka, 2009, Cashflow Risk, Systematic Earnings Revisions, and the Cross-Section of Stock Returns, *Journal of Financial Economics* 94, 448-468.

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

Denis, D. K., McConnell, J. J., Ovtchinnikov, A. V., Yu, Y., 2003, S&P 500 index additions and earnings expectations, *Journal of Finance* 58, 1821-1840.

Dhillon, U., and Johnson, H., 1991, Changes in the Standard and Poor's 500 list, *Journal of Business*, 64, 75-85.

Dimson, E., 1979, Risk measurement when shares are subject to infrequent trading, *Journal of Financial Economics* 7, 1821-1840.

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.

Ellul, Andrew, Pab Jotikasthira, and Christian T. Lundblad, 2010, Regulatory Pressure and Fire Sales in the Corporate Bond Market, Indiana University.

Evans, Martin, and Richard Lyons, 2002, Order Flow and Exchange Rate Dynamics, *Journal of Political Economy*, 110, pp. 170-180.

Fama, Eugene F. and Kenneth R. French, 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, Eugene F. and Kenneth R. French, 1996, Multifactor Explanations of Asset Pricing Anomalies, *Journal of Finance* 51, 55–84.

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

Froot, Kenneth and Emil Dabora, 1999, How Are Stock Prices Affected by the Location of Trade? *Journal of Financial Economics*, 53, pp. 189-216.

Fung, William and David A. Hsieh, 2001, The Risk in Hedge Fund Strategies: Theory and Evidence from Trend Followers, *Review of Financial Studies* 14, 313-341.

Fung, William and David A. Hsieh, 2004, Hedge fund benchmarks: A risk-based approach, *Financial Analysts Journal* 60, 65-80.

Getmansky, Mila, Andrew W. Lo, and Igor Makarov, 2004, An econometric model of serial correlation and illiquidity in hedge fund returns, *Journal of Financial Economics* 74, 529-609.

Giglio, Stefano, 2011, Credit Default Swap Spreads and Systemic Financial Risk, Harvard University working paper.

Green, T. Clifton, and Byoung-Hyoun Hwang, 2009, Price-Based Return Co-movement, *Journal of Financial Economics*, 93:1, pp. 37-50.

Greenwood, Robin, 2005, A Cross Sectional Analysis of the Excess Comovement of Stock Returns, working paper, Harvard Business School, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=698162.

Greenwood, Robin and Nathan Sosner, 2007, Trading Patterns and Excess Comovement of Stock Returns, *Financial Analysts Journal*, 63:5, pp. 69-81.

Greenwood, Robin and David Thesmar, 2009, Stock Price Fragility, *Journal of Financial Economics*, forthcoming.

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

Gromb, Denis and Dimitri Vayanos, 2010, Limits of Arbitrage: State of the Theory, *Annual Review of Financial Economics*, forthcoming.

Harford, Jarrad and Aditya Kaul, 2009, Correlated Order Flow: Pervasiveness, Sources, and Pricing Effects, *Journal of Financial and Quantitative Analysis*, forthcoming.

Harris, L., and E. Gurel, 1986, Price and Volume Effects Associated with Changes in the S&P 500: new evidence for the existence of price pressure, *Journal of Finance* 41, 851-860.

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

Jegadeesh, Narasimhan and Sheridan Titman, 1993, Returns to Buying and Selling Losers: Implications for Stock Market Efficiency, *Journal of Finance* 48, 65–91.

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

Koch, Andy, Stefan Ruenzi, and Laura Starks, 2009, Commonality in Liquidity: A Demand-Side Explanation, working paper, University of Texas.

Lee, Charles, and Mark Ready, 1991, Inferring Trade Direction from Intraday Data, *Journal of Finance*, 46, pp. 733-746.

Longstaff, Francis, Sanjay Mithal, and Eric Neis, 2005, Corporate Yield Spreads: Default Risk or Liquidity? New Evidence from the Credit Default Swap Market, *Journal of Finance* 60:5, pp. 2213-2253.

Lou, Dong, 2009, A Flow-Based Explanation for Return Predictability, working paper, London School of Economics.

Mayordomo, Sergio, Ignacio Peña, and Eduardo Schwartz, 2010, Are All Credit Default Swap Databases Equal?, NBER Working Paper No. w16590.

Merton, Robert C., 1980, 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.

Pindyck, Richard and Julio Rotemberg, 1993, The Comovement of Stock Prices, *Quarterly Journal of Economics*, 108:4, 1073-1104.

Roll, Richard, Eduardo Schwartz, and Avanidhar Subrahmanyam, 2009, Liquidity and the Law of One Price: The Case of the Futures/Cash Basis, *Journal of Finance*, 62:5, pp. 2201-2234.

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.

Schwert, G. William, 2003, Anomalies and Market Efficiency, *Handbook of the Economics of Finance*, Chapter 15, 937-972, eds. George Constantinides, Milton Harris, and Rene M. Stulz, North-Holland.

Shleifer, A., 1986, Do Demand Curves for Stocks Slope Down?, *Journal of Finance* 41, 579-590.

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, 2004, 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.

Vijh, Anand, 1994, 'S&P 500 Trading Strategies and Stock Betas,' *Review of Financial Studies*, 7:1, pp. 215-251.

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

Wurgler, J., and Zhuravskaya, K., 2002, Does Arbitrage flatten Demand Curves for Stocks? *Journal of Business* 75, 583-608.