Yale ICF Working Paper No. 08-08

Comovement

Nicholas Barberis
Yale University

Andrei Shleifer
Harvard University

Jeffrey Wurgler
New York University

October 2003

Comovement

Nicholas Barberis, Andrei Shleifer and Jeffrey Wurgler
University of Chicago, Harvard University and New York University

October 2003*

Abstract

We consider two broad views of return comovement: the traditional view, derived from frictionless economies with rational investors, which attributes it to comovement in news about fundamental value, and an alternative view, in which market frictions or noise-trader sentiment delink it from comovement in fundamentals. Building on Vijh (1994), we use data on inclusions into the S&P 500 to distinguish these views. After inclusion, a stock's beta with the S&P goes up. In bivariate regressions which control for the return of non-S&P stocks, the increase in S&P beta is even larger. These results are generally stronger in more recent data. Our findings cannot easily be explained by the fundamentals-based view and provide new evidence in support of the alternative friction- or sentiment-based view.

*We thank John Campbell, Ned Elton, Kenneth French, Will Goetzmann, Joel Hasbrouck, Steven Kaplan, Anthony Lynch, Mike Ryngaert, Bill Schwert, Robert Shiller, Tuomo Vuolteenaho, two anonymous referees, seminar participants at Columbia University, Harvard University, the London School of Economics, New York University, Rice University, UCLA, the University of Florida at Gainesville, Yale University and the NBER for helpful comments, Huafeng Chen and Bill Zhang for helpful comments and outstanding research assistance, the Q Group for financial support, and Rick Mendenhall and Standard and Poor's for providing data.

1 Introduction

Researchers have uncovered numerous patterns of comovement in asset returns. There are strong common factors in the returns of small-cap stocks, value stocks, closed-end funds, stocks in the same industry and bonds of the same rating and maturity. There is common movement of individual stocks within national markets and also among international markets themselves.

A substantial body of work tries to understand whether the sensitivity of asset returns to common factors such as these can help explain *average* rates of return. Much less work, however, has been done to understand why common factors arise in the first place. Why do certain groups of assets comove while others do not? What determines assets' betas on these common factors? In this paper, we consider two broad theories of comovement – one traditional, the other more novel – and present new evidence to distinguish them.

The traditional theory, derived from economies without frictions and with rational investors, holds that comovement in prices reflects comovement in fundamental values. In a frictionless economy with rational investors, price equals fundamental value – an asset's rationally forecasted cash flows discounted at a rate appropriate for their risk – and so any comovement in prices must be due to comovement in fundamentals.

In economies with frictions or with irrational investors, and in which there are limits to arbitrage, comovement in prices is delinked from comovement in fundamentals. This suggests a second broad class of "friction-based" and "sentiment-based" theories of comovement. We examine three specific views of comovement that can be described in these terms.

The first is the category view, analyzed by Barberis and Shleifer (2003). They argue that, to simplify portfolio decisions, many investors first group assets into categories such as small-cap stocks, oil industry stocks, or junk bonds, and then allocate funds at the level of these categories rather than at the individual asset level.[1] If some of the investors using categories are noise traders with correlated sentiment, and if their trading affects prices, then as they move funds from one category to another, their coordinated demand induces common factors in the returns of assets that happen to be classified into the same category, even when these assets' cash flows are uncorrelated.

Another kind of comovement, which we refer to as the habitat view, starts from the observation that many investors choose to trade only a subset of all available securities. Such preferred habitats may arise because of transaction costs, international trading restrictions, or lack of information. As these investors' risk aversion, sentiment or liquidity needs change, they alter their exposure to the securities in their habitat, thereby inducing a common

[1] See Mullainathan (2001) for a more general analysis of the use of categories for processing information.

factor in the returns of these securities. This view of comovement predicts that there will be a common factor in the returns of securities that are held and traded by a specific subset of investors, such as individual investors.[2]

A third view of comovement, the information diffusion view, holds that, due to some market friction, information is incorporated more quickly into prices of some stocks than others. For example, some stocks may be less costly to trade, or may be held by investors with faster access to breaking news and the resources required to exploit it. In this view, there will be a common factor in the returns of stocks that incorporate information at similar rates: when good news about aggregate earnings is released, some stocks reflect it today and move up together immediately; the remaining stocks also move up together, but only after some delay.[3]

Early evidence of friction- or sentiment-based comovement can be found in Vijh (1994), who studies changes in the market betas of stocks that are added to the S&P 500 index. Standard and Poor's emphasizes that in choosing stocks for inclusion into the S&P 500, they are simply trying to make their index as representative as possible of the overall U.S. economy, not signaling a view about fundamental value. If this is the case, inclusion should not change investors' perception of the correlation between the stock's fundamental value and other stocks' fundamental values. Under the traditional view of comovement, then, inclusion should not change the correlation of the included stock's return with the returns of other stocks.

The friction- and sentiment-based views make a different prediction. Consider first the category and habitat views. The vast popularity of S&P-linked investment products such as S&P mutual funds, futures and options, suggests that the index is a preferred habitat for some investors and a natural category for many more. When a stock is added to the S&P, it enters a category (habitat) used by many investors and is buffeted by fund flows in and out of that category (habitat). If arbitrage is limited, these fund flows raise the correlation of the included stock's return with the returns of other stocks in the S&P.

The information diffusion view also predicts a rise in correlation between the added stock's return and the S&P return. Under this view, stocks in the S&P 500 are quick to incorporate news about aggregate cash flows, perhaps because they have particularly low trading costs or are held by investors with better access to news. When a stock enters the S&P, it starts to incorporate market-wide news at the same time as other S&P stocks, rather

[2]Other models that consider investor habitats are motivated by similar informational and transaction cost considerations as our own, but focus on different issues. Merton (1987) analyzes the cross-sectional implications when investors apply standard mean-variance analysis, but only over a subset of available assets. Our focus is on the effects of habitat-level demand shifts that affect all stocks in the habitat equally.

[3]A related view, that makes many similar predictions, is that it is market-wide sentiment and not just market-wide cash-flow news that is incorporated more quickly into some stocks than into others. In this view, there will be a common factor in the returns of stocks that incorporate sentiment at similar rates.

than a day later, say. As a result, it comoves more with other S&P stocks after inclusion than before.

In fact, Vijh (1994) finds that at both daily and weekly frequencies, stocks added to the index between 1975 and 1989 experience a significant increase in their betas on the value-weighted return on NYSE and AMEX stocks, a close proxy for the value-weighted S&P return. In contrast to the fundamentals view, but consistent with the three friction- or sentiment-based views, addition to the index does lead to a shift in the correlation structure of returns.[4]

In this paper, we return to inclusions into the S&P 500 and provide new evidence in support of the friction- or sentiment-based theories of comovement. First, by applying a univariate regression analysis similar to Vijh's (1994) to the longer sample of data now available, we uncover considerably larger effects. While stocks added to the S&P during 1976-1987 experience an average increase in daily S&P beta of 0.067 after inclusion, the average increase in the 1988-2000 period is 0.214. In light of the growing importance of the S&P 500 as both a category and a habitat, the fact that the effect is not only present but *larger* in more recent data is especially supportive of the friction- or sentiment-based views.

Our second and principal contribution is to introduce a *bivariate* regression test to distinguish the two broad theories of comovement. The friction- and sentiment-based views imply that, in a bivariate regression of a stock's return on both the S&P and a non-S&P "rest of the market" index, the S&P beta should go up after inclusion while the non-S&P beta should fall; that these patterns should go in the opposite direction for stocks deleted from the index; and that they should be stronger in more recent data. The fundamentals view, in contrast, predicts no shift in S&P and non-S&P betas after inclusion.

The bivariate regressions provide evidence of friction- or sentiment-based comovement altogether stronger than that uncovered by the univariate tests. At daily, weekly and even monthly frequencies, the bivariate tests show a striking increase in S&P beta and decline in non-S&P beta. At the daily frequency, for example, S&P beta increases by 0.326 on average over the 1976-2000 period, while non-S&P beta falls by 0.319. Significant results in the opposite direction are observed when stocks are deleted from the index, and our effects are quantitatively larger in more recent data.

We examine the robustness of these findings in several ways. We obtain similar results when analyzing the data from a "calendar time" rather than an "event time" perspective. We also find that the results cannot be attributed to thin trading. Finally, we show that the shifts in S&P betas for stocks added to the index are much larger than those for a

[4] Vijh (1994) computes betas with the overall market return rather than with the S&P return because his main objective is to understand whether investor trading patterns affect the standard measure of asset risk – market beta – and not to design the most powerful test of trading-based comovement.

sample of "matched" stocks, namely stocks similar to the event stocks on a number of characteristics, but which are *not* included into the index. This last observation rules out certain fundamentals-based explanations of our results under which inclusion into the S&P *coincides* with a change in the correlations between stocks' fundamental values, even if it does not cause such a change.

While our results are in principle consistent with all three variants of friction- or sentiment-based comovement – category, habitat and information diffusion – we also attempt to determine how big a role each of them plays. In particular, we decompose the shift in betas around inclusion into one component due to information diffusion and a residual component, more likely due to category and habitat effects. The distinguishing feature of the information diffusion story is that it predicts non-zero cross-autocorrelations between S&P and non-S&P returns, making it possible to identify its effect by including lead and lag terms in the regressions. Our analysis suggests that some small fraction of the daily univariate results and a much larger fraction of the daily bivariate results are indeed due to information diffusion effects.

Our paper adds to a growing body of evidence that can be naturally understood as friction- or sentiment-based comovement. Froot and Dabora (1999) show that the returns of Royal Dutch shares are surprisingly delinked from the returns of Shell shares, even though the two stocks are claims to the same cash-flow stream and therefore have the same fundamental value. Hardouvelis, La Porta and Wizman (1994) and Bodurtha, Kim and Lee (1995) show that closed-end country funds comove as much with the stock market in the country where they are traded as with the stock market in the country where their *assets* are traded. Lee, Shleifer and Thaler (1991) find that domestic closed-end funds often comove with small-cap stocks even when their asset holdings consist of large-cap stocks. Pindyck and Rotemberg (1990) uncover evidence of excess comovement in the prices of seven major commodities. Fama and French (1995) show that it is hard to connect the strong common factors in the returns of value stocks and small stocks to common factors in news about earnings.[5]

Greenwood and Sosner (2002) test the friction- and sentiment-based views using data on additions to and deletions from the Nikkei index. They find increases in beta and R^2 following addition to the index, and decreases following deletions. Their evidence is therefore also consistent with our predictions; if anything, the results for the Japanese data are even stronger than those for the U.S. data.

A large literature examines whether index inclusion has a contemporaneous effect on

[5]In principle, fundamentals-based comovement can be generated through news about discount rates, as well as through news about earnings. However, changes in interest rates or risk aversion induce a common factor in the returns on *all* stocks, and do not explain why a particular subset of stocks comove. A common factor in news about the risk of certain assets may also be a source of comovement for those assets, but there is little evidence to support such a mechanism in the case of small stocks or value stocks.

price levels. Harris and Gurel (1986), Shleifer (1986), Lynch and Mendenhall (1997) and Wurgler and Zhuravskaya (2002) find strong price effects for S&P 500 inclusions, while Kaul, Mehrotra and Morck (2000) and Greenwood (2002) find similar effects in the Toronto Stock Exchange TSE 300 and Nikkei 225 indices, respectively. This literature argues that uninformed demand affects price levels; our paper shows that it may also affect patterns of comovement.[6]

In Section 2, we illustrate the basic predictions of our three friction- or sentiment-based views of comovement. In Section 3, we test these predictions using data on S&P 500 inclusions and deletions. Section 4 concludes.

2 Theories of Comovement

The traditional theory of return comovement is the fundamentals-based view, under which the returns of two assets are correlated because changes in the assets' fundamental values are correlated. In this section, we briefly present reduced-form models of our three alternative friction- or sentiment-based theories of comovement: the category, habitat and information diffusion views. The models yield the predictions that motivate the empirical work in Section 3.

Consider an economy that contains a riskless asset in perfectly elastic supply and with zero rate of return, and also $2n$ risky assets in fixed supply. Risky asset i is a claim to a single liquidating dividend $D_{i,T}$ to be paid at some later time T. This eventual dividend equals

$$D_{i,T} = D_{i,0} + \varepsilon_{i,1} + \ldots + \varepsilon_{i,T}, \tag{1}$$

where $D_{i,0}$ and $\varepsilon_{i,t}$ are announced at time 0 and time t, respectively, and where

$$\varepsilon_t = (\varepsilon_{1,t}, \ldots, \varepsilon_{2n,t})' \sim N(0, \Sigma_D), \text{ i.i.d over time.}$$

The price of a share of risky asset i at time t is $P_{i,t}$ and the return on the asset between time $t-1$ and time t is

$$\Delta P_{i,t} \equiv P_{i,t} - P_{i,t-1}. \tag{2}$$

For simplicity, we refer to the asset's change in price as its return.

[6]Numerous other papers also present evidence consistent with uninformed demand affecting prices. These include French and Roll (1986), Goetzmann and Massa (2001), Gompers and Metrick (2001), Cooper, Dimitrov and Rau (2001), Mitchell, Pulvino and Stafford (2002) and Lamont and Thaler (2003). Cooper, Dimitrov and Rau (2001) is particularly related to our category view in that they show that stocks that recategorized themselves as dot-com companies in the late 1990s experienced large price increases even when their business model did not change.

Suppose that to simplify their decision-making, some investors group the $2n$ risky assets into two categories, X and Y, and then allocate funds at the level of these categories rather than at the individual asset level. In particular, they place assets 1 through n in category X and assets $n+1$ through $2n$ in category Y. It may be helpful to think of X and Y as "old economy" and "new economy" stocks, respectively.

Suppose now that these categories are also adopted by noise traders, who channel funds in and out of the categories depending on their sentiment. A simple representation for asset returns is then

$$\Delta P_{i,t} = \varepsilon_{i,t} + \Delta u_{X,t}, \quad i \epsilon X \tag{3}$$

$$\Delta P_{j,t} = \varepsilon_{j,t} + \Delta u_{Y,t}, \quad j \epsilon Y, \tag{4}$$

where

$$\begin{pmatrix} u_{X,t} \\ u_{Y,t} \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \sigma_u^2 \begin{pmatrix} 1 & \rho_u \\ \rho_u & 1 \end{pmatrix} \right), \text{ i.i.d. over time.} \tag{5}$$

Here, $u_{X,t}$ can be thought of as time t noise trader sentiment about the securities in category X. Since the noise traders allocate funds by category, this sentiment level is the same for all securities in category X. Equation (3) says that the return on a security in category X is affected not only by news about cash flows, but also by changes in sentiment about X: when noise traders become more bullish about old economy stocks, these stocks go up in price.

Equations (3) and (4) can also be thought of as a reduced form for the habitat view of comovement. In this case, X and Y simply have to be reinterpreted as habitats, not categories: instead of representing groups of assets that some investors do not distinguish between when allocating funds, they represent groups of assets that are the sole holdings of some investors. Specifically, we can think of assets 1 through n as U.S. stocks and assets $n+1$ through $2n$ as U.K. stocks; there are many investors in both countries who restrict themselves to trading only domestic securities. Under the habitat interpretation, $u_{X,t}$ can be thought of as a variable tracking the risk aversion, sentiment or liquidity needs of investors who invest only in the securities in X. The returns of assets in habitat X are affected not only by news about cash flows but also by changes in the risk aversion, say, of these specific groups of investors.

The third variant of friction- or sentiment-based comovement, the information diffusion view, can be modelled as:

$$\Delta P_{i,t} = \varepsilon_{i,t}, \; i \epsilon X \tag{6}$$

$$\Delta P_{j,t} = \mu \varepsilon_{j,t} + (1-\mu)\varepsilon_{j,t-1}, \; j \epsilon Y. \tag{7}$$

Here, X and Y are groups of stocks that, for some reason, incorporate new information at different rates. Stocks in X incorporate news announced at time t immediately. Stocks in

Y only incorporate a fraction μ of time t news immediately; the remaining fraction, $1-\mu$, is incorporated in the following period.

For the fund flows of noise traders or investors with preferred habitats to affect prices, as in equations (3)-(4), or for information to be incorporated into stock prices with delay, as in equations (6)-(7), arbitrage must be limited in some way, perhaps because arbitrageurs have short horizons (De Long et al. 1990, Shleifer and Vishny 1997). The idea that there are limits to arbitrage has found support in the considerable empirical literature, cited in the introduction, suggesting that demand unrelated to news about fundamental value affects security prices.[7]

To uncover evidence of friction- or sentiment-based comovement, we look for testable predictions that emerge from our reduced-form models. One set of predictions describes what happens when a stock moves from one category or habitat to another, or from a group of stocks that incorporates information slowly to one that does so quickly. Such reclassification occurs in many ways. For example, if the market capitalization of a large-cap stock declines sufficiently, it enters the small-cap stock category. More simply, stocks are regularly added to categories like the S&P 500 to replace stocks that have been removed due to bankruptcies or mergers.

In the Appendix, we show that under quite general conditions, the reduced-form models (3)-(4) and (6)-(7) yield the following prediction:

Prediction 1: Suppose that risky asset j, previously a member of Y, is reclassified into X and that the cash-flow covariance matrix Σ_D is constant over time. Then, as the number of risky assets $n \to \infty$, the probability limit of the OLS estimate of β_j in the univariate regression

$$\Delta P_{j,t} = \alpha_j + \beta_j \Delta P_{X,t} + v_{j,t}, \tag{8}$$

where

$$\Delta P_{X,t} = \frac{1}{n} \sum_{l \epsilon X} \Delta P_{l,t}, \tag{9}$$

as well as the probability limit of the R^2 of this regression, increase after reclassification.

The intuition is straightforward, whether X and Y are categories, habitats or groups of stocks that incorporate information at similar rates. To take the category view, when asset j enters category X, it is buffeted by noise traders' flows of funds in and out of that category. This increases its covariance with the return on category X, $\Delta P_{X,t}$, and hence also its beta loading on that return. For simplicity, we assume a fixed cash-flow covariance matrix in the

[7] Barberis, Shleifer and Wurgler (2002) show more formally that in an economy where rational arbitrageurs interact either with category-based noise traders or with investors with preferred habitats, returns have the form given in equations (3)-(4).

statement of the prediction. A more general version would predict that beta increases *more* than can be explained by any change in cash-flow correlations.

A similar intuition lies behind the next prediction which, as we show in the Appendix, can also be derived from equations (3)-(4) and (6)-(7):

Prediction 2: Suppose that risky asset j, previously a member of Y, is reclassified into X, and that the cash-flow covariance matrix Σ_D is fixed. Then, as the number of risky assets $n \to \infty$, the probability limit of the OLS estimate of $\beta_{j,X}$ in the bivariate regression

$$\Delta P_{j,t} = \alpha_j + \beta_{j,X} \Delta P_{X,t} + \beta_{j,Y} \Delta P_{Y,t} + v_{j,t} \tag{10}$$

rises after reclassification, while the probability limit of the OLS estimate of $\beta_{j,Y}$ falls. In particular, the pre-reclassification values of the two slope coefficients, $\underline{\beta}_{j,X}$ and $\underline{\beta}_{j,Y}$, and their post-reclassification values, $\overline{\beta}_{j,X}$ and $\overline{\beta}_{j,Y}$, satisfy

$$\begin{aligned} \underline{\beta}_{j,X} &= 0, \ \underline{\beta}_{j,Y} = 1 \\ 0 &< \overline{\beta}_{j,X} < 1, \ 0 < \overline{\beta}_{j,Y} < 1, \ \overline{\beta}_{j,X} + \overline{\beta}_{j,Y} = 1. \end{aligned}$$

The intuition for this prediction is again straightforward, however X and Y are interpreted. Consider again the category view, whose basic prediction is that when a stock enters category X, it becomes more sensitive to the category X sentiment shock $\Delta u_{X,t}$. The independent variable in the Prediction 1 regression, $\Delta P_{X,t}$, is not a clean measure of this sentiment shock: as equation (3) shows, a substantial part of its variation comes from news about cash flows. The $\Delta P_{Y,t}$ variable in regression (10) can be thought of as a control for such news, making the coefficient on $\Delta P_{X,t}$ a cleaner measure of sensitivity to $\Delta u_{X,t}$. Alternatively, under the information diffusion view, regression (10) is a cleaner test than regression (8) of whether, after inclusion, stock j becomes more sensitive to that component of market-wide news that is incorporated more quickly into X than into Y. Note that while $\beta_{j,X}$ rises after reclassification, it rises by less than 1, and that while $\beta_{j,Y}$ falls, it falls by less than 1. Moreover, the rise in $\beta_{j,X}$ has the same absolute magnitude as the fall in $\beta_{j,Y}$.

Later on, when we try to distinguish not only between the fundamentals-based and the friction- or sentiment-based theories of comovement, but also between the three specific variants of friction- or sentiment-based comovement, we make use of a prediction of the information diffusion view not shared by the category and habitat views, namely that there will be positive cross-autocorrelation between S&P and non-S&P returns. Under the information diffusion view, news about aggregate cash flows is reflected in S&P stocks today, but only with some delay in non-S&P stocks.

The three friction- or sentiment-based views we consider depend on the existence of noise traders who allocate funds by category, on the existence of distinct investor habitats, or on

heterogeneous rates of information diffusion. If these features are not present, Predictions 1 and 2 will not hold. In this case,

$$\begin{aligned} \Delta P_{i,t} &= \varepsilon_{i,t}, \ i\epsilon X \\ \Delta P_{j,t} &= \varepsilon_{j,t}, \ j\epsilon Y, \end{aligned} \tag{11}$$

and comovement is entirely fundamentals-based, in that the correlation of returns is completely determined by the correlation of cash-flow news. If, as assumed in the predictions, the cash-flow covariance matrix Σ_D remains constant, the correlation structure of returns also remains constant. In other words, β_j and R^2 in Prediction 1 and $\beta_{j,X}$ and $\beta_{j,Y}$ in Prediction 2 remain unchanged after reclassification.

3 Empirical Tests

To test Predictions 1 and 2, we need a group of securities with three characteristics. First, the group must be a natural category or preferred habitat for many investors, or else a set of stocks that incorporate information more quickly than do other stocks. Second, since the predictions concern reclassification, there must be clear and identifiable changes in group membership. Finally, to control for fundamentals-based comovement, a security's inclusion or removal from the group should not change investors' perception of the correlation between the security's fundamental value and the fundamental values of other securities in the group.

Stocks in the S&P 500 index satisfy our three requirements. The vast popularity of S&P-linked investment products such as S&P mutual funds, futures and options, suggests that the index is a preferred habitat for some investors and a natural category for many more. Even if category-based or habitat-based investors trade S&P futures and options rather than the underlying stocks, any influence they have on the prices of these futures and options is quickly transmitted to the cash market by index arbitrageurs. Moreover, the high liquidity of S&P stocks and the fact that they are traded by sophisticated financial institutions makes it likely that they will be particularly quick to incorporate market-wide cash-flow news.

The S&P 500 also has the second characteristic we require: there is clear and identifiable turnover in its membership. In a typical year there are about 30 changes; our full sample, which we describe in Section 3.1., contains 455 additions and 76 deletions.

Finally, the act of adding a stock to the S&P 500 should not change investors' perceptions of the covariance between the stock's fundamental value and other stocks' fundamental values. The stated goal of Standard and Poor's is to make the index representative of the U.S. economy, not to signal a view about future cash flows.[8] Deletions from the index are

[8]Denis, McConnell, Ovtchinnikov and Lu (2003) find that index additions coincide with increases in

another matter. Stocks are usually removed from the index because a firm is merging, being taken over, or nearing bankruptcy. In these situations cash-flow characteristics *are* changing, so we exclude these cases from our deletion sample.

In Sections 3.2. and 3.3., we test Predictions 1 and 2 for the case where X is the S&P 500 and Y is stocks outside that index. Our null hypothesis is that return comovement is entirely a function of comovement in news about fundamentals, so that the betas and R^2 in regressions (8) and (10) do *not* change after inclusion or deletion. The alternative hypothesis is that return comovement is delinked from fundamentals as described by the category, habitat and information diffusion views, so that the betas and R^2 change as stated in the predictions. In Sections 3.4. through 3.6., we examine the robustness of our findings. In Section 3.7., we attempt to decompose any deviations from fundamentals-based comovement that we uncover across the three specific friction- or sentiment-based views.

3.1 Data

We study S&P 500 index inclusions between September 22, 1976 and December 31, 2000 and deletions between January 1, 1979 and December 31, 2000. Standard & Poor's did not record announcement dates of index changes before September 1976 and we are unable to obtain data on deletions before 1979.

There are 590 inclusion events in the inclusion sample period and 565 deletions in the deletion sample period. Inclusion events are excluded if the new 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 is not available. This last condition binds only in the case of monthly data, for which we need a longer post-event sample – in this case, events after December 31, 1998 are discarded. The final sample covers 455 inclusions in the case of daily and weekly data, and 324 in the case of monthly data.

Deletion events are excluded if the firm is involved in a merger, takeover, or bankruptcy proceeding, or if required return data is not available. These circumstances, determined by searching the NEXIS database, exclude the vast majority of deletions. The ones that remain are largely cases where stocks were deleted because the index became too heavily weighted in the industry they belonged to. The final sample contains 76 deletions in the case of daily and weekly data; lack of a long enough post-event sample for events after December 31, 1998

earnings. These results would not seem to explain the growth in the S&P inclusion effect over time, nor the large price effects observed by Kaul, Mehrotra and Morck (2000) and Greenwood (2002) for completely mechanical reweightings of stocks that are *already* members of indices. 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.

reduce this to just 45 deletions in the case of monthly data.[9]

3.2 Univariate Regressions

Prediction 1 holds that under the three friction- or sentiment-based views we consider, stocks added to (deleted from) the S&P 500 will comove more (less) with the other members of the index after the addition (deletion) event.

To test this, for each inclusion and deletion, we estimate the univariate regression

$$R_{j,t} = \alpha_j + \beta_j R_{SP500,t} + v_{j,t} \tag{12}$$

separately for the period before the event and the period after the event, and record the change in the slope coefficient, $\Delta\beta_j$, and the change in the R^2, ΔR_j^2. $R_{j,t}$ is the return of the event stock between time $t-1$ and t, while $R_{SP500,t}$ is the contemporaneous return on the S&P 500 index, obtained from the CRSP Index on the S&P Universe file.[10]

We run these regressions for three data frequencies: daily, weekly and monthly. With daily and weekly data, the pre-event regression is run over the 12-month period ending the month before the month of the inclusion *announcement*, while the post-event regression is run over the 12-month period starting the month after the month of the inclusion *implementation*. With monthly data, we extend the estimation period to 36 months.[11]

Table 1 reports the change in slope coefficient, averaged across all events in the sample, $\overline{\Delta\beta}$, as well as the average change in R^2, $\overline{\Delta R^2}$. It confirms that stocks added to the S&P 500 experience a strongly significant increase in daily and weekly betas and R^2. In the full sample of additions, the mean increase in daily beta is 0.151 and in weekly data, 0.110. At the monthly frequency, however, neither the 0.042 increase in beta nor the slight increase in R^2 are significant at conventional levels. Other than a weakly significant change in daily beta, we do not detect significant drops in beta or R^2 around deletion events.

Our univariate regression results are similar to those of Vijh (1994), who studies inclusions into the S&P during 1975-1989. Using daily and weekly data, Vijh (1994) computes changes

[9] The S&P 500 inclusion and deletion data are available on request.

[10] To avoid spurious effects, we remove the contribution of the stock in question from the right-hand side variable. For addition events, this means that there are 500 stocks in the right-hand side variable before the addition, and 499 afterward. The reverse applies for deletion events.

[11] Up until October 1989, inclusions and deletions were made effective on the day of their announcement. Since then, the changes have been announced a week in advance of their implementation. It is not clear whether to view the to-be-added stock as being in the index or not in the index during the time between announcement and implementation; significant price effects have been documented on both days (Lynch and Mendenhall, 1997). To avoid these issues entirely, we do not use data from the month of the announcement or of the implementation; these are usually the same month.

in individual stocks' betas with the value-weighted return of all NYSE and AMEX stocks. He finds average increases in market beta of 0.08 and 0.037 at the daily and weekly frequencies, respectively. These numbers are comparable in magnitude to the 0.067 and 0.025 increases in S&P beta we report for our 1976-1987 subsample.

Our evidence shows that in the more recent data now available, the results are considerably stronger than those in the 1975-1989 sample alone: Table 1 shows that at daily and weekly frequencies, the increases in beta and R^2 across inclusion events are quantitatively larger in the second subsample. At the daily frequency, for example, the increase in S&P beta is 0.067 in the first subsample, but rises to 0.214 in the second. The difference is highly statistically significant. Given the growing importance of the S&P 500 as an investment class, the fact that the effect is not only present but larger in more recent data is especially supportive of the category and habitat views. This evidence can also be consistent with the information diffusion view if, because of changes in transaction costs or ownership, S&P stocks now incorporate information even more quickly, relative to non-S&P stocks, than they did before.

The three friction- or sentiment-based views also predict that the shifts in beta after inclusion should become weaker at sufficiently low frequencies. Since we expect noise trader sentiment to revert *eventually*, and even slowly-diffusing information to be incorporated *eventually*, lower frequency returns, and therefore also lower frequency patterns of comovement, will be more closely tied to fundamentals. The shifts in beta after inclusion need not, however, decline monotonically with the horizon over which returns are measured. Under the category view, for example, if sentiment shocks are positively autocorrelated at the first few lags and negatively autocorrelated only thereafter, the shifts in beta should first increase with the return horizon and only later, decline.[12]

The univariate regressions in Table 1 conform to the simplest prediction of the three friction- or sentiment-based views: as we move from the daily results to the monthly ones, the shifts in beta decline in both economic and statistical significance. This suggests either that any noise-trader sentiment reverts within a matter of days; or that information, while not immediately reflected in prices, is nonetheless incorporated quite quickly; or both. In Section 3.7., we will be able to say more as to which possibility is most likely.

The standard errors deserve comment. If two events are close together in calendar time, there may be substantial overlap in the time periods covered by the regressions associated with each event. This means that the disturbances $v_{j,t}$ may be correlated across events, which in turn implies that the $\{\Delta\beta_j\}_{j=1,\ldots,N}$ may not be independent but rather autocorrelated at several lags.

[12] In the model of equations (3)-(4), the shifts in beta after inclusion do decline monotonically with the return horizon: the simplifying assumption of i.i.d sentiment means that changes in sentiment are negatively autocorrelated even at the first lag.

We use simulation methods to compute standard errors that account for this dependence. A full description of our methodology can be found in the Appendix. In brief, we generate a simulated data set, consisting of an S&P return and returns on included stocks, and set the cross-sectional correlation of the disturbance terms to whatever value generates a first-order autocorrelation in the $\Delta\beta_j$'s equal to that observed in our results. We then compute $\overline{\Delta\beta}$ in this sample, under the null that betas do *not* change after inclusion. By generating many such data sets, we obtain the distribution of $\overline{\Delta\beta}$ under the null, and hence also, appropriate standard errors.[13]

3.3 Bivariate Regressions

Our primary empirical contribution is to introduce a bivariate regression test to distinguish the fundamentals-based theory of comovement from the friction- or sentiment-based theories. This test, presented in Prediction 2, holds that under the friction- or sentiment-based views, controlling for the return of *non*-S&P stocks, a stock that is added to (removed from) the S&P will experience a large increase (decrease) in its loading on the S&P return. To test this, for each inclusion and deletion, we run the bivariate regression

$$R_{j,t} = \alpha_j + \beta_{j,SP500} R_{SP500,t} + \beta_{j,\text{non}SP500} R_{\text{non}SP500,t} + v_{j,t} \tag{13}$$

for the period before the event and the period after the event, and record the changes in S&P and non-S&P betas, $\Delta\beta_{j,SP500}$ and $\Delta\beta_{j,\text{non}SP500}$. $R_{\text{non}SP500,t}$ is the return on non-S&P stocks in the NYSE, AMEX and Nasdaq universe between time $t-1$ and time t. This is inferred from index return and capitalization data using the identity that the capitalization-weighted average return of S&P stocks and of non-S&P stocks equals the overall CRSP value-weighted return on NYSE, AMEX and Nasdaq stocks. We run the regressions at daily, weekly and monthly frequencies and use the same 12-month and 36-month estimation windows as for the univariate tests.

Table 1 reports the change in S&P beta, averaged across all events in the sample, $\overline{\Delta\beta}_{SP500}$, as well as the average change in non-S&P beta, $\overline{\Delta\beta}_{\text{non}SP500}$. These bivariate regression results are statistically stronger than the univariate regression results. At all three data frequencies, S&P 500 inclusion is associated with a substantial and significant increase in beta with the S&P and a substantial and significant decrease in beta with the rest of the market. For example, daily beta with the S&P 500 goes up by an average of 0.326 and average daily beta with other stocks drops by -0.319. Large and significant results also obtain for deletion events at the daily and weekly frequencies. Moreover, the table shows that at all

[13] At least for daily and weekly frequencies, cross-correlation of disturbances does not affect the standard errors by very much. The reason is that such cross-correlation produces positive autocorrelation in the $\Delta\beta_j$'s at the first few lags but negative autocorrelation at higher lags. As a result, the variance of $\overline{\Delta\beta}$ is only slightly higher than if the disturbances were uncorrelated.

three frequencies, the changes in S&P betas are quantitatively larger in the second subsample, although the differences relative to the first subsample are not statistically significant.

Not only are the shifts in betas around inclusion and deletion events statistically significant, but by one metric at least, they are economically significant as well. One way of judging this is the following. For fixed values of the right-hand side variables in regression (13), we can generate predicted values of the left-hand side variable, first using betas estimated before inclusion and then betas estimated after inclusion. In results available on request, we find that the two sets of predicted values are quite different – the standard deviation of their difference is relatively large – providing one demonstration that the shift in betas is economically meaningful.

Unlike the univariate regression results, the bivariate regressions do not display the other pattern predicted by the three friction- or sentiment-based views of comovement, namely that the results should be stronger at higher frequencies. One possible explanation is discussed in Section 3.6 below. There, we show that when we control more carefully for fundamentals-based comovement, the results are closer to the predicted pattern. In brief, we find that in the bivariate tests, fundamentals-based comovement makes up a larger fraction of the monthly beta shifts than of the daily beta shifts. Once this fundamentals-based component is taken out, the residual effect, which we attribute to friction- or sentiment-based comovement, is indeed smaller at lower frequencies.

A noteworthy feature of the bivariate regression (13) is collinearity. The usual standard errors do, of course, take the correlation of the explanatory variables into account – no special correction is required. Collinearity does mean that the standard errors on the slope coefficients in the bivariate regressions are likely to be higher than on the slope coefficient in the univariate regressions. Nonetheless, Table 1 shows that despite the larger standard errors, the bivariate regressions reject the null more strongly than do the univariate tests.

Collinearity is, however, at the root of another feature of the bivariate regression results, namely that the changes in S&P and non-S&P betas are of similar but opposite magnitude. To see the link, recall that when two right-hand side variables are highly correlated, the sum of their respective slope coefficients is estimated much more accurately than either of the individual coefficients (Goldberger 1979, p.251).[14] In the context of regression (13), the sum of $\beta_{j,SP500}$ and $\beta_{j,\text{non}SP500}$ is estimated more precisely than either individual slope, which immediately implies that the sum of $\overline{\Delta\beta}_{SP500}$ and $\overline{\Delta\beta}_{\text{non}SP500}$ is also estimated more precisely than either of the two component pieces. However, the sum of $\overline{\Delta\beta}_{SP500}$ and $\overline{\Delta\beta}_{\text{non}SP500}$ is also approximately the average change in overall *market* beta, which we know from Vijh (1994) to be close to zero. It is therefore no surprise that the changes in S&P and non-S&P betas are of similar but opposite magnitudes.

[14] Mathematically, this is because collinearity induces negative correlation between the two slope estimates, lowering the variance of their sum.

The fact that average changes in S&P and non-S&P betas are roughly equal and opposite has little relevance for our test of the friction- and sentiment-based views of comovement: this pattern is to be expected both under the null of fundamentals-based comovement *and* under the friction- or sentiment-based alternatives. The key distinction is that under the null, changes in S&P and non-S&P beta should not be statistically different from zero, while under the alternative, S&P betas should exhibit a significant *increase* after inclusion, and non-S&P betas a significant *decrease*. The bivariate regression results in Table 1 provide clear support for the alternative.

3.4 Calendar Time Tests

The methodology used in Table 1 is an "event time" approach. An alternative technique, the "calendar time" approach, is often used to address a common statistical problem in event studies, namely correlation of returns across events. As described in Section 3.2., we use simulations to deal with this problem. Calendar time tests offer a second way of checking that our results are robust to this statistical consideration.

The calendar time approach requires the construction of two portfolios: a "pre-event" portfolio whose return at time t, $R_{pre,t}$, is the equal-weighted average return at time t of all stocks that *will* be added to the index within some window after time t; and a "post-event" portfolio whose return at time t, $R_{post,t}$, is the equal-weighted average return at time t of all stocks that *have been* added to the index within some window preceding time t. For daily and weekly data, we take the window to be a year, and extend it to three years for monthly data.

The calendar time test of Prediction 1 calls for running two regressions,

$$R_{pre,t} = \alpha_{pre} + \beta_{pre} R_{SP500,t} + v_{pre,t} \tag{14}$$

and

$$R_{post,t} = \alpha_{post} + \beta_{post} R_{SP500,t} + v_{post,t}, \tag{15}$$

and checking whether $\beta_{post} > \beta_{pre}$ and whether the R^2 in the second regression is greater than in the first.

Similarly, the calendar time test of Prediction 2 calls for running the two regressions

$$R_{pre,t} = \alpha_{pre} + \beta_{pre,SP500} R_{SP500,t} + \beta_{pre,\text{non}SP500} R_{\text{non}SP500,t} + v_{pre,t} \tag{16}$$

and

$$R_{post,t} = \alpha_{post} + \beta_{post,SP500} R_{SP500,t} + \beta_{post,\text{non}SP500} R_{\text{non}SP500,t} + v_{post,t}, \tag{17}$$

and checking whether $\beta_{post,SP500} > \beta_{pre,SP500}$ and $\beta_{post,\text{non}SP500} < \beta_{pre,\text{non}SP500}$.

Table 2 reports the changes in slope coefficients and R^2s. The results are again supportive of the friction- or sentiment-based views. In the univariate regressions, significant increases in beta and R^2 occur at the daily and weekly frequencies, and for R^2, even at the monthly frequency. In the bivariate regressions, the results for inclusion events are strongly significant at all three data frequencies, although the results for deletion events are weaker than in the event time tests, turning up no statistically significant effects.

3.5 Alternative Explanations: Characteristic and Industry Effects

We now consider some alternative explanations for the results in Table 1. One possibility is that stocks in the S&P 500 index differ from other stocks in terms of some characteristic, and that the stocks Standard and Poor's chooses to include are those increasingly demonstrating that characteristic. If the characteristic is also associated with a cash-flow factor, this may explain our results.

The most obvious such characteristic is size. Stocks in the S&P have considerably higher market capitalizations than stocks outside the index, and the stocks Standard and Poor's includes into the index have often been growing in size prior to inclusion. Moreover, size is known to be associated with a cash-flow factor: there is a common component to news about the earnings of large-cap stocks (Fama and French 1995). Our finding that S&P betas increase around inclusion may then arise because included stocks are growing in size around inclusion and are therefore increasingly loading on the large stock cash-flow factor. More generally, this is a story in which inclusion into the S&P *coincides* with a change in the cash-flow covariance matrix, even if it does not cause it.

A related concern is based on industry effects. Suppose that some industry becomes increasingly dominant in the economy. This increases the fraction of the value of the S&P made up by stocks in this industry. Moreover, to keep their index representative, Standard and Poor's may start drawing an increasing number of new inclusions from this industry. Since S&P beta is computed using the *value*-weighted S&P return, this simultaneity can also lead to effects like those we observe: if Yahoo is included into the S&P at precisely the time that other technology stocks in the index are growing in value – as indeed it was, having been added in December 1999 – it may covary more with the S&P after inclusion than before.

We address both these competing explanations with a matching exercise. For each event stock included into the S&P during our sample period, we search for a "matching" stock, drawn from the same industry as the event stock and in the same size decile both at the time of inclusion and 12 months before inclusion, but which is *not* included into the index. Since the matching stock matches the event stock on industry and on recent growth in market capitalization, it is arguably as good a candidate for inclusion as the event stock itself, but

simply happens not to be included. If the matching stocks do not demonstrate the same increase (decrease) in S&P (non-S&P) beta as the event stocks, it strengthens the case that the results in Table 1 are due to friction- or sentiment-based comovement, rather than to the alternative explanations. In the case of deleted stocks, the matching stock is a stock in the S&P which matches the deleted stock on industry and recent change in market capitalization, but which is not removed from the index.[15]

Figure I and Table 3 contain the results of the matching exercise. Panels A, B and C of the figure present results for daily, weekly and monthly data, respectively. Within each panel, the top two graphs present results for the event firms, while the bottom two correspond to the matching firms. Also, within each panel, the graphs on the left present results for the first half of our sample, while those on the right correspond to the second half.

The graphs use rolling regressions to show how S&P and non-S&P betas change in event time. Consider the top left graph in Panel A. The solid line shows the mean daily S&P beta and the dashed line shows the mean daily non-S&P beta, re-estimated each month using the prior 12 months of daily data. Coefficients plotted to the left of the left vertical line therefore use only pre-event returns, while coefficients plotted to the right of the right vertical line use only post-event returns. Coefficients in between use both pre- and post-event data. To be clear, the steady change in estimated betas between the two vertical lines should not be interpreted as a steady change in true betas. Rather, it arises from mixing data from the pre- and post-event regimes.[16]

Figure I indicates that whichever frequency we look at, the matching stocks exhibit much smaller shifts in betas than do the event stocks. Table 3, which reports the changes in betas and R^2s in univariate and bivariate regressions for event stocks relative to the analogous changes for matching stocks, confirms this impression. At the daily and weekly frequencies, the changes in betas and R^2s in univariate regressions and in S&P and non-S&P betas in bivariate regressions remain strongly significant across inclusion events, even net of the changes for matching stocks. At the monthly frequency, the results are weaker than those

[15] At the monthly frequency, in order to match the window betas are computed over, we look for matching stocks that match the event stock on size both at inclusion and 36 months before inclusion. At all frequencies, we initially try to match by SIC4 industry code. If no match can be found, we allow the matching stock to be in the same SIC3 industry class, then to be within one size decile at inclusion, then to be within one size decile 12 months before inclusion, then to be in the same SIC2 industry class, then to be within two size deciles at inclusion, then to be within two size deciles 12 months before inclusion, and finally to be within three size deciles 12 months before inclusion. Events for which no such matches can be found are not included in the matching exercise samples.

[16] In terms of these graphs, the beta changes reported in Table 1 are the average beta as of event month +12, which uses data from months [+1, +12] minus the average beta as of event month -1, which uses data from months [-12, -1]. There are fewer data points in the graph ($N = 169$) than in Table 1 ($N = 196$ for additions in the first subsample), however, because the graph includes only event firms with available return data for a full 24 months after inclusion and for which we are able to find matching firms.

in Table 1, but are still highly significant in the second subsample.[17]

In Section 3.3., we noted another prediction of the three friction- or sentiment-based views, namely that the beta shifts around inclusion should decline in significance at lower frequencies. Table 3 shows that once we control more carefully for fundamentals-based comovement in the bivariate regressions, the residual effect, which is more likely to reflect friction- or sentiment-based comovement, does indeed decline at lower frequencies. The increase in S&P beta, for example, falls from 0.318 at the daily frequency to 0.173 at the monthly one. In Section 3.7., we will be able to say more as to which of the three alternative views is most responsible for this declining effect.

At the same time, our matching exercise disrupts the declining pattern of betas seen in the univariate regressions in Table 1. Table 3 shows that the average change in beta in the univariate regressions declines from 0.120 to 0.077 as we go from the daily to the weekly frequency, but then increases to 0.104 at the monthly frequency. Overall, though, accounting for fundamentals-based comovement appears to improve the fit of our results with the declining significance of beta shifts predicted by the friction- or sentiment-based views.

3.6 Alternative Explanations: Non-trading Effects

Standard and Poor's explicitly selects stocks that are highly liquid and frequently traded for inclusion into the S&P 500 index. Nevertheless, microstructure research suggests that our daily frequency results might still have some spurious upward bias due to non-trading effects (Scholes and Williams 1977, Dimson 1979).

To see this, suppose that towards the end of the trading day, some positive market-wide information is announced. Since stocks in the S&P are heavily traded, it is probable that they will trade again before the end of the day: the S&P return for that day, $R_{SP500,t}$, is therefore likely to reflect the good news. A stock *outside* the index, however, typically trades less often, and may not trade again before the end of the day. As a result, its return for that day, $R_{j,t}$, may *not* reflect the good news. A regression of $R_{j,t}$ on $R_{SP500,t}$ then produces an artificially low slope coefficient. Once the stock is included into the S&P 500 though, non-trading ceases to be a problem and the slope coefficient in the regression goes up, just as it does in our results.[18]

[17] In Table 1, we conducted simulations to correct the standard errors for possible correlation in disturbance terms across regressions. This problem affects matching stock regressions just as much as it does event stock regressions, but it does not affect *differences* in slopes across the two sets of regressions. The Table 3 standard errors are therefore the usual ones – no simulation-based correction is required.

[18] Even though stocks added to the S&P may not trade as often, prior to their inclusion, as do stocks already in the index, they invariably still trade at least once a day. So while non-trading may be partly

Under the non-trading hypothesis, then, the beta increases we observe occur simply because the typical added stock trades more frequently after inclusion. To see if our results are indeed due to non-trading, we adopt a test suggested by Vijh (1994). We split the sample of included stocks into two groups: those whose turnover decreases after addition and those whose turnover increases. More precisely, for each stock, we compute the average monthly turnover (share volume divided by shares outstanding) over the same pre- and post-event windows used to compute betas, and assign the stock to the first group if its post-event average turnover is lower than pre-event average turnover, and to the second group otherwise.

If non-trading is driving our results, we should only see increases in S&P beta after inclusion for the second group of stocks – those whose turnover increases – while for the first group, we should see decreases.[19] The three friction- or sentiment-based views of comovement, on the other hand, predict increases in S&P beta for *both* groups. Table 4 presents the results. It shows that in both the univariate and bivariate regressions, there are strongly statistically significant increases in S&P betas even for stocks whose turnover decreases after inclusion. Non-trading cannot therefore be the primary driver of our daily frequency results.

The results in Table 4 do, however, suggest some role for non-trading. In the univariate regressions, the shifts in S&P beta are considerably larger in Panel B, which is what we would expect if non-trading played some part in our results. In the bivariate regressions, the ratios of the absolute increase in S&P beta to the absolute decrease in non-S&P beta are larger in Panel B, which is again what would be expected in the presence of some non-trading. We can think of the quantitative effect of friction- or sentiment-based comovement on betas, controlling for non-trading, as some average of the results in Panels A and B – somewhere between 0.083 and 0.194 in the univariate case, and somewhere between 0.313 and 0.350 in the bivariate case. Clearly, these effects will be strongly statistically significant and economically substantial.

In summary, non-trading effects take a small bite out of the daily frequency results in Table 1. Characteristic and industry effects, on the other hand, reduce the strength of the monthly frequency results in Table 1, but have little impact on the daily and weekly findings. Taken together, Tables 3 and 4 show that even after adjusting for both non-trading and matching firm effects, there is still strong residual evidence for friction- or sentiment-based comovement. The daily and weekly results remain strongly significant, as do the monthly results in the second subsample.

---

behind the daily results in Table 1, it cannot plausibly explain the weekly and monthly results.

[19] Strictly speaking, we should see increases in S&P beta for those included stocks whose turnover moves closer to the value-weighted turnover of stocks already in the S&P. Since, as Vijh (1994) shows, almost all included stocks have turnover lower than this value-weighted average, both before and after inclusion, the prediction is more simply made in terms of an increase in turnover.

3.7 Distinguishing the Category, Habitat and Information Diffusion Views

So far, we have argued that a substantial part of the shift in betas around inclusion is due to friction- or sentiment-based comovement of some kind, but have not identified which of the three mechanisms – category, habitat, or information diffusion – might be playing a more significant role. In this section, we attempt to shed some light on this.

Using return data alone, it is difficult to distinguish the category and habitat views, but it may still be possible to isolate the effect of information diffusion. The reason is that the information diffusion view makes a prediction not shared by the category and habitat views, namely that there should be positive cross-autocorrelation between S&P and non-S&P returns: good news about aggregate earnings is incorporated into the prices of S&P stocks today, but only with some delay into the prices of non-S&P stocks.

This suggests that we can identify the effects of information diffusion by including leading and lagged S&P and non-S&P returns in the univariate and bivariate regressions, (12) and (13). Specifically, for daily frequency data, we estimate the regressions

$$R_{j,t} = \alpha_j + \sum_{s=-5}^{5} \beta_j^{(s)} R_{SP500,t+s} + v_{j,t} \tag{18}$$

and

$$R_{j,t} = \alpha_j + \sum_{s=-5}^{5} \beta_{j,SP500}^{(s)} R_{SP500,t+s} + \sum_{s=-5}^{5} \beta_{j,\text{non}SP500}^{(s)} R_{\text{non}SP500,t+s} + v_{j,t} \tag{19}$$

both before and after each inclusion and deletion event, thereby including five leads and lags. The changes in the quantities

$$\sum_{s=-5}^{5} \beta_j^{(s)}, \ \sum_{s=-5}^{5} \beta_{j,SP500}^{(s)}, \text{ and } \sum_{s=-5}^{5} \beta_{j,\text{non}SP500}^{(s)} \tag{20}$$

across event dates can be interpreted as the beta shifts that would occur in a world without any delay in the incorporation of information: in other words, they capture the component of the results in Table 1 due only to category and habitat effects and not to information diffusion. If they are closer to zero than the daily beta shifts in Table 1, then we know that information diffusion is in part driving our results. If they are still non-zero, however, we know that information diffusion cannot explain the entire shift in betas, but that category and habitat effects are also at work.

Panel A of Table 5 reports the average changes in the quantities in (20) across event dates. Comparing these numbers to those in Table 1, we see that information diffusion accounts for about a third of the economic magnitude of our univariate results and that even after removing its effects, the residual, attributable to category and habitat effects,

remains strongly significant. It also shows that information diffusion accounts for a much larger fraction, perhaps two-thirds, of the economic magnitude of the bivariate regression results. Even though the increase in S&P beta remains quantitatively large, particularly in the second subsample, the standard errors are so large that statistical significance drops considerably.[20]

Panel B shows, for the full sample of inclusions, how the components of the three quantities in (20) change after inclusion. For example, the row labelled t reports the changes in $\beta_j^{(0)}$, $\beta_{j,SP500}^{(0)}$ and $\beta_{j,\text{non}SP500}^{(0)}$; the row labelled $t-1$ reports the changes in $\beta_j^{(-1)}$, $\beta_{j,SP500}^{(-1)}$ and $\beta_{j,\text{non}SP500}^{(-1)}$; and so on for the other rows.

These numbers are consistent with information diffusion playing some role in our results. The loading of the added stock's return on the previous day's S&P return declines after inclusion – $\beta_j^{(-1)}$ and $\beta_{j,SP500}^{(-1)}$ fall by 0.080 and 0.118 respectively – consistent with the idea that before inclusion, the stock incorporates market-wide information a day after the S&P, but as soon as it is included, it does so the same day. The rise, after inclusion, in the sensitivity of the added stock's return to the following day's non-S&P return can be interpreted the same way.

Panel B shows that the pattern predicted by the information diffusion view – a drop in the sensitivity, $\beta_j^{(-s)}$ and $\beta_{j,SP500}^{(-s)}$, of the added stock's return to lagged S&P returns – is barely detectable beyond just a one day lag in the univariate regressions and beyond a two day lag in the bivariate regressions. This suggests that although information does appear to enter non-S&P returns more slowly than S&P returns, the delay is short.[21]

In our discussion of Table 3, we noted that the shifts in beta across inclusion are weaker at lower frequencies. This pattern disappears after taking out the effect of information diffusion: the 0.109 and 0.104 univariate and bivariate shifts in *daily* S&P betas, reported in Table 5, are not statistically different from our best estimates of the univariate and bivariate shifts in *monthly* S&P beta, reported in Table 3 as 0.104 and 0.173. The frequency effect in Table 3 is therefore probably driven mostly by information diffusion: if information can take two or three days to be fully incorporated, there will be a more pronounced change in beta at the daily frequency than at weekly or monthly frequencies. The residual shift in beta, after controlling for information diffusion, is relatively stable across frequencies. Since this residual is most likely attributable to the category and habitat views, it suggests that

[20] The component of our results that we attribute to slow information diffusion could also be due to a closely related mechanism, in which it is market-wide sentiment, and not just market-wide cash-flow news, that is incorporated more quickly into some stocks than into others.

[21] The rapid attenuation of the information diffusion effect in Panel B suggests that it can play only a minimal role at lower frequencies. A back-of-the-envelope calculation indicates that information diffusion should affect the weekly univariate results in Table 1 by approximately $\frac{1}{5}\sum_{s=-5}^{-1}(-s)\beta_j^{(s)} + \frac{1}{5}\sum_{s=1}^{5}(s)\beta_j^{(s)}$, with an analogous formula holding for the bivariate regressions; in either case, this number is very small.

the noise-trader sentiment assumed by those mechanisms does not mean-revert quickly, but rather is quite persistent.

4 Conclusion

We present two broad theories of return comovement and examine them empirically using data on inclusions into the S&P 500. The traditional theory, fundamentals-based comovement, is derived from frictionless economies with rational investors and attributes comovement in returns to correlation in news about fundamental value. The alternative theory argues that, due to market frictions or noise-trader sentiment, return comovement is delinked from fundamentals. We consider three specific kinds of such "friction-based" or "sentiment-based" comovement, and label them the category, habitat and information diffusion views.

In earlier work on S&P inclusions during 1975-1989, Vijh (1994) finds that stocks added to the index experience a significant increase in beta after inclusion. In this paper, we first show that by applying a univariate analysis similar to Vijh's (1994) to the new data now available, we uncover considerably stronger effects. We then show that in *bivariate* regressions that also include the return of non-S&P stocks, the rise in S&P beta is altogether larger than anything in the univariate regressions, that this last effect is also stronger in more recent data and that exactly the opposite pattern holds for stocks deleted from the index. We also find that in both univariate and bivariate regressions, the effects are somewhat weaker at lower frequencies.

Our findings cannot easily be explained by the fundamentals-based view of comovement, but fit with the friction- or sentiment-based views. We make some progress in determining the relative importance of the three specific mechanisms we propose. Slow diffusion of information, for example, appears to account for a relatively small part of the beta shifts in the univariate regressions, but for up to two-thirds of the beta shifts in the daily bivariate regressions. Overall, our evidence adds to the growing range of phenomena identified by financial economists that reveal the importance for valuation of market frictions and noise-trader sentiment.

5 Appendix

Derivation of Predictions 1 and 2

For simplicity, we assume that the cash-flow covariance matrix Σ_D takes a specific form, although the predictions also hold for more general structures. In particular, we suppose that the cash-flow shock to an asset has three components: a market-wide cash-flow shock, a group-specific cash-flow shock which affects assets in one group but not the other, and a completely idiosyncratic cash-flow shock specific to the asset. Formally, for $i \epsilon X$,

$$\varepsilon_{i,t} = \psi_M f_{M,t} + \psi_S f_{X,t} + \sqrt{(1 - \psi_M^2 - \psi_S^2)} f_{i,t}, \tag{21}$$

and for $j \epsilon Y$,

$$\varepsilon_{j,t} = \psi_M f_{M,t} + \psi_S f_{Y,t} + \sqrt{(1 - \psi_M^2 - \psi_S^2)} f_{j,t}, \tag{22}$$

where $f_{M,t}$ is the market-wide shock, $f_{X,t}$ and $f_{Y,t}$ are the group-specific shocks, and $f_{i,t}$ and $f_{j,t}$ are idiosyncratic shocks; ψ_M and ψ_S are constants that control the relative importance of the three components. Each shock has unit variance and is orthogonal to the other shocks.

Consider first the reduced-form model for the category and habitat views, equations (3)-(4). Suppose that asset $n+1$ is reclassified from Y into X, and that at the same moment, asset 1 is reclassified from X into Y. Before reclassification,

$$\begin{aligned} \Delta P_{X,t} &= \varepsilon_{X,t} + \Delta u_{X,t} \\ \Delta P_{Y,t} &= \varepsilon_{Y,t} + \Delta u_{Y,t} \\ \Delta P_{n+1,t} &= \varepsilon_{n+1,t} + \Delta u_{Y,t}, \end{aligned} \tag{23}$$

where

$$\varepsilon_{k,t} = \frac{1}{n} \sum_{l \epsilon k} \varepsilon_{l,t},\ k = X, Y, \tag{24}$$

while after reclassification, $\Delta P_{X,t}$ and $\Delta P_{Y,t}$ are still given by (23), but now

$$\Delta P_{n+1,t} = \varepsilon_{n+1,t} + \Delta u_{X,t}. \tag{25}$$

Using the standard formula for regression coefficients, we find that as $n \to \infty$, the probability limit of the OLS estimate of β_{n+1} in the regression

$$\Delta P_{n+1,t} = \alpha_{n+1} + \beta_{n+1} \Delta P_{X,t} + v_{n+1,t} \tag{26}$$

is given by

$$\beta_{n+1} = \frac{\psi_M^2 + 2\sigma_u^2 \rho_u}{\psi_M^2 + \psi_S^2 + 2\sigma_u^2}$$

before reclassification, and by

$$\beta_{n+1} = \frac{\psi_M^2 + 2\sigma_u^2}{\psi_M^2 + \psi_S^2 + 2\sigma_u^2}$$

afterwards, thereby confirming that β_{n+1} increases after reclassification as claimed in Prediction 1. Moreover, since $\text{var}(\Delta P_{n+1,t})$ and $\text{var}(\Delta P_{X,t})$ are unchanged after reclassification, the increase in β_{n+1} also implies an increase in the R^2 of regression (26) after inclusion.

Similarly, we find that as $n \to \infty$, the probability limit of the OLS estimates of $\beta_{n+1,X}$ and $\beta_{n+1,Y}$ in the regression

$$\Delta P_{n+1,t+1} = \alpha_{n+1} + \beta_{n+1,X}\Delta P_{X,t+1} + \beta_{n+1,Y}\Delta P_{Y,t+1} + v_{n+1,t+1} \tag{27}$$

are given by

$$\beta_{n+1,X} = 0,\ \beta_{n+1,Y} = 1 \tag{28}$$

before reclassification and by

$$\beta_{n+1,X} = \frac{2\sigma_u^2(1-\rho_u)}{\psi_S^2 + 2\sigma_u^2(1-\rho_u)},\ \beta_{n+1,Y} = \frac{\psi_S^2}{\psi_S^2 + 2\sigma_u^2(1-\rho_u)} \tag{29}$$

afterwards, thereby confirming Prediction 2.

Now consider the reduced-form model for the information diffusion view, equations (6)-(7). Suppose that asset $n+1$ is reclassified from Y into X, and that at the same moment, asset 1 is reclassified from X into Y. Before reclassification,

$$\begin{aligned} \Delta P_{X,t} &= \varepsilon_{X,t} \\ \Delta P_{Y,t} &= \mu\varepsilon_{Y,t} + (1-\mu)\varepsilon_{Y,t-1} \\ \Delta P_{n+1,t} &= \mu\varepsilon_{n+1,t} + (1-\mu)\varepsilon_{n+1,t-1}, \end{aligned} \tag{30}$$

where $\varepsilon_{k,t}$ is defined in (24), while after reclassification, $\Delta P_{X,t}$ and $\Delta P_{Y,t}$ are still given by (30), but now

$$\Delta P_{n+1,t} = \varepsilon_{n+1,t}. \tag{31}$$

This implies that as $n \to \infty$, the probability limit of the OLS estimate of β_{n+1} in regression (26) is given by

$$\beta_{n+1} = \frac{\mu\psi_M^2}{\psi_M^2 + \psi_S^2} \tag{32}$$

before reclassification, and by

$$\beta_{n+1} = \frac{\psi_M^2}{\psi_M^2 + \psi_S^2} \tag{33}$$

afterwards, thereby confirming that β_{n+1} increases after reclassification as claimed in Prediction 1. Moreover, since $\text{var}(\Delta P_{n+1,t})$ and $\text{var}(\Delta P_{X,t})$ are unchanged after reclassification, the increase in β_{n+1} also implies an increase in the R^2 of regression (26) after inclusion.

Similarly, we find that as $n \to \infty$, the probability limit of the OLS estimates of $\beta_{n+1,X}$ and $\beta_{n+1,Y}$ in regression (27) are given by

$$\beta_{n+1,X} = 0,\ \beta_{n+1,Y} = 1 \tag{34}$$

before reclassification and by

$$\begin{aligned}\beta_{n+1,X} &= \frac{(1-\mu)^2\psi_M^2(\psi_M^2+\psi_S^2)}{\mu^2\psi_S^2(2\psi_M^2+\psi_S^2)+(1-\mu)^2(\psi_M^2+\psi_S^2)^2}, \\ \beta_{n+1,Y} &= \frac{\mu\psi_S^2(2\psi_M^2+\psi_S^2)}{\mu^2\psi_S^2(2\psi_M^2+\psi_S^2)+(1-\mu)^2(\psi_M^2+\psi_S^2)^2}\end{aligned} \tag{35}$$

afterwards. It is straightforward to check that if $\mu < \mu^*$, where

$$\mu^* = \frac{(\psi_M^2+\psi_S^2)^2}{\psi_S^2(2\psi_M^2+\psi_S^2)+(\psi_M^2+\psi_S^2)^2}, \tag{36}$$

then $\beta_{n+1,X}$ increases after inclusion, while $\beta_{n+1,Y}$ falls. If, moreover, $\mu = \mu^{**}$, where

$$\mu^{**} = \frac{\psi_M^2+\psi_S^2}{3\psi_M^2+2\psi_S^2}, \tag{37}$$

then the increase in $\beta_{n+1,X}$ is precisely equal in magnitude to the decrease in $\beta_{n+1,Y}$. Prediction 2 does not therefore follow as generally from equations (6)-(7) as it does from equations (3)-(4), but it is nonetheless derivable from both reduced form models.

Standard error computations

As discussed in Section 3.2., correlation in the residuals $v_{j,t}$ across different stocks means that the $\{\Delta\beta_j\}_{j=1,\ldots,N}$ are not independent, thereby requiring an adjustment to the standard error of the test statistic $\overline{\Delta\beta}$ that would be obtained assuming independence. To gauge, at least approximately, how large this correction should be, we generate a large number of artificial data sets and use them to compare the distribution of the test statistic in the case where there *is* correlation in residuals to the distribution in the case where there is none.

We simulate the artificial data as follows. We generate the S&P and non-S&P returns from

$$\begin{aligned}R_{SP500,t} &= R_{M,t} + w_{SP500,t} \\ R_{\text{non}SP500,t} &= R_{M,t} + w_{\text{non}SP500,t}\end{aligned} \tag{38}$$

where

$$\begin{aligned}R_{M,t} &\sim N(0,\sigma_M^2) \\ w_{SP500,t} &\sim N(0,\sigma_S^2) \\ w_{\text{non}SP500,t} &\sim N(0,\sigma_S^2),\end{aligned} \tag{39}$$

all three variables assumed i.i.d over time, and independent of all other variables. We generate the included stock's return from

$$R_{j,t} = R_{\text{non}SP500,t} + R_{F,t} + w_{j,t} \tag{40}$$

where

$$\begin{aligned} R_{F,t} &\sim N(0, \sigma_F^2) \\ w_{j,t} &\sim N(0, \sigma^2 - \sigma_M^2 - \sigma_S^2 - \sigma_F^2), \end{aligned} \tag{41}$$

both variables assumed i.i.d over time and independent of all other variables. $R_{F,t}$ is a common factor in the returns of included stocks that leads to correlation in the disturbances $v_{j,t}$ across regressions. The variance of $w_{j,t}$ in (41) ensures that the overall variance of the included stock's return in (40) is given by σ^2.

We set σ_M to 0.2 to match the standard deviation of the market return; σ_S to 0.07 to generate a correlation between $R_{SP500,t}$ and $R_{\text{non}SP500,t}$ that matches the empirical value of 0.85; and σ to 0.5, to match the typical standard deviation of an individual stock. The one parameter that remains is σ_F, and we set it at whatever value generates a first-order autocorrelation in the $\{\Delta\beta_j\}_{j=1,\ldots,N}$ equal to that in the actual data. These autocorrelations are typically around 0.1, a level which can be generated by setting $\sigma_F = 0.15$.

With parameter values assigned, we now generate many artificial data sets. Each data set consists of an S&P return and non-S&P return spanning the entire length of the sample and, taking the case of daily data, return data for 455 included stocks for the two years around their respective event dates. The spacing of event dates in the artificial data matches the spacing in the actual data. With a large number of artificial data sets in hand, we can construct the distribution of the test statistic $\overline{\Delta\beta}$ for the case where the $\{\Delta\beta_j\}_{j=1,\ldots,N}$ are not independent. We then compare this to the distribution of the test statistic for the case of $\sigma_F = 0$, in which the $\{\Delta\beta_j\}_{j=1,\ldots,N}$ are uncorrelated. Finally, we inflate the standard errors in the tables by the implied correction factor.

6 References

Barberis, N., and A. Shleifer (2003), "Style Investing," *Journal of Financial Economics* 68, 161-199.

Barberis, N., Shleifer A., and J. Wurgler (2001), "Comovement," NBER Working Paper no. 8895.

Bodurtha J., Kim, D., and C.M. Lee (1995), "Closed-end Country Funds and U.S. Market Sentiment," *Review of Financial Studies* 8, 879-918.

Cooper, M., Dimitrov, O., and P.R. Rau (2001), "A Rose.com By Any Other Name," *Journal of Finance* 56, 2371-2388.

De Long, J.B., Shleifer A., Summers L., and R. Waldmann (1990), "Noise Trader Risk in Financial Markets," *Journal of Political Economy* 98, 703-38.

Denis D.K., McConnell, J.J., Ovtchinnikov, A.V., and Y. Lu (2003), "S&P 500 Index Additions and Earnings Expectations," *Journal of Finance* 58, 1821-1840.

Dimson, E. (1979), "Risk Measurement When Shares are Subject to Infrequent Trading," *Journal of Financial Economics* 7, 197-227.

Fama, E., and K. French (1995), "Size and Book-to-Market Factors in Earnings and Returns," *Journal of Finance* 50, 131-155.

French, K., and R. Roll (1986), "Stock Return Variances: The Arrival of Information and the Reaction of Traders," *Journal of Financial Economics* 17, 5-26.

Froot, K., and E. Dabora (1999), "How are Stock Prices affected by the Location of Trade?," *Journal of Financial Economics* 53, 189-216.

Goetzmann, W., and M. Massa (2003), "Index Funds and Stock Market Growth," forthcoming, *Journal of Business*.

Goldberger, A. (1991), *A Course in Econometrics*, Harvard University Press: Cambridge, MA.

Gompers, P., and A. Metrick (2001), "Institutional Investors and Equity Prices," *Quarterly Journal of Economics* 116, 229-259.

Greenwood, R. (2002), "Short- and Long-term Demand Curves for Stocks: Some Evidence from Japan," Working Paper, Harvard University.

Greenwood, R., and N. Sosner (2002), "Where Do Betas Come From?," Working Paper, Harvard University.

Hardouvelis, G., La Porta R., and T. Wizman (1994), "What Moves the Discount on Country Equity Funds?," in Jeffrey Frankel (ed.), *The Internationalization of Equity Markets*, Chicago, The University of Chicago Press.

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.

Kaul, A., Mehrotra V., and R. Morck (2000), "Demand Curves for Stocks Do Slope Down: New Evidence from an Index Weights Adjustment," *Journal of Finance* 55, 893-912.

Lamont, O., and R. Thaler (2003), "Can the Market Add and Subtract? Mispricing in Tech Stock Carveouts," *Journal of Political Economy* 111, 227-268.

Lee, C., Shleifer A., and R. Thaler (1991), "Investor Sentiment and the Closed-end Fund puzzle," *Journal of Finance* 46, 75-110.

Lynch, A. and R. Mendenhall (1997), "New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index," *Journal of Business* 70, 351-83.

Merton, R. (1987), "A Simple Model of Capital Market Equilibrium with Incomplete Information," *Journal of Finance* 42, 483-510.

Mitchell, M., Pulvino, T., and E. Stafford (2002), "Limited Arbitrage in Equity Markets," *Journal of Finance* 57, 551-584.

Mullainathan, S. (2001), "Thinking Through Categories," Working paper, MIT.

Pindyck, R., and J. Rotemberg (1990), "The Excess Comovement of Commodity Prices," *Economic Journal* 100, 1173-1189.

Scholes M. and J. Williams (1977), "Esimating Betas from Nonsynchronous Data," *Journal of Financial Economics* 5, 309-327

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

Shleifer, A., and R. Vishny (1997), "The Limits of Arbitrage," *Journal of Finance* 52, 35-55.

Vijh, A. (1994), "S&P 500 Trading Strategies and Stock Betas," *Review of Financial Studies*

7, 215-251.

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

Table 1. Changes in comovement of stocks added to and deleted from the S&P 500 Index. Changes in the slope and the fit of regressions of returns of stocks added to and deleted from the S&P 500 Index on returns of the S&P 500 Index and the non-S&P 500 rest of the market. The sample includes stocks added to and deleted from the S&P 500 between 1976 and 2000 which were not involved in mergers or related events (described in the text), and which have sufficient return data on CRSP. For each added stock *j*, the univariate model

$$R_{j,t} = \alpha_j + \beta_j R_{SP500,t} + \upsilon_{j,t}$$

and the bivariate model

$$R_{j,t} = \alpha_j + \beta_{j,SP500} R_{SP500,t} + \beta_{j,nonSP500} R_{nonSP500,t} + \upsilon_{j,t}$$

are separately estimated for the pre-change and post-change period. Returns on the S&P 500 (R_{SP500}) are from the CRSP Index on the S&P 500 Universe file. Returns on a capitalization-weighted index of the non-S&P 500 stocks ($R_{nonSP500}$) in the NYSE, AMEX, and Nasdaq are inferred from the identity

$$R_{VWCRSP,t} = \left(\frac{CAP_{CRSP,t-1} - CAP_{SP500,t-1}}{CAP_{CRSP,t-1}} \right) R_{nonSP500,t} + \left(\frac{CAP_{SP500,t-1}}{CAP_{CRSP,t-1}} \right) R_{SP500,t} \,.$$

Total capitalization on the S&P 500 (CAP_{SP500}) is from the CRSP Index on the S&P 500 Universe file. Returns on the value-weighted CRSP NYSE, AMEX, and Nasdaq index (R_{VWCRSP}) and total capitalization (CAP_{CRSP}) are from the CRSP Stock Index file. Returns from October 1987 are excluded. The mechanical influence of the added or deleted stock is removed from the independent variables as appropriate. For the univariate regression model, we examine the mean difference between the pre-change slope and the post-change slope $\overline{\Delta\beta}$, and the mean change in fit $\overline{\Delta R^2}$. For the bivariate model, we examine the mean changes in the slopes, $\overline{\Delta\beta_{SP500}}$ and $\overline{\Delta\beta_{nonSP500}}$. The pre-change and post-change estimation periods are [-12,-1] and [+1,+12] months for daily and weekly returns and [-36,-1] and [+1,+36] months for monthly returns. Panels A, B, and C show results for daily, weekly, and monthly returns, respectively. Standard errors are determined by simulation, to account for cross-correlation, and are reported in parentheses. ***, **, and * denote significant differences from zero at the 1%, 5%, and 10% levels in one-sided tests, respectively.

Sample		N	Univariate		Bivariate	
			$\overline{\Delta\beta}$ (s.e.)	$\overline{\Delta R^2}$ (s.e.)	$\overline{\Delta\beta_{SP500}}$ (s.e.)	$\overline{\Delta\beta_{nonSP500}}$ (s.e.)
Panel A. Daily Returns						
Additions	1976-2000	455	0.151*** (0.021)	0.049*** (0.005)	0.326*** (0.027)	-0.319*** (0.033)
	1976-1987	196	0.067*** (0.023)	0. 038*** (0.008)	0.286*** (0.041)	-0.301*** (0.050)
	1988-2000	259	0.214*** (0.032)	0.058*** (0.007)	0.357*** (0.036)	-0.332*** (0.045)
Deletions	1979-2000	76	-0.087* (0.049)	-0.010 (0.007)	-0.511*** (0.111)	0.550*** (0.122)
Panel B. Weekly Returns						
Additions	1976-2000	455	0.110*** (0.029)	0.033*** (0.008)	0.174** (0.053)	-0.119** (0.056)
	1976-1987	196	0.025 (0.036)	0.027** (0.012)	0.137 (0.094)	-0.125 (0.093)
	1988-2000	259	0.173*** (0.043)	0.037*** (0.010)	0.202*** (0.061)	-0.115* (0.069)
Deletions	1979-2000	76	-0.129 (0.105)	-0.015 (0.010)	-0.505*** (0.161)	0.412** (0.169)
Panel C. Monthly Returns						
Additions	1976-1998	324	0.042 (0.033)	0.004 (0.011)	0.317*** (0.069)	-0.252*** (0.064)
	1976-1987	172	-0.010 (0.044)	0.006 (0.015)	0.267*** (0.100)	-0.167* (0.091)
	1988-1998	152	0.101** (0.051)	0.000 (0.015)	0.375*** (0.093)	-0.348*** (0.088)
Deletions	1979-1998	45	0.006 (0.097)	0.001 (0.021)	0.303 (0.240)	-0.256 (0.252)

Table 2. Changes in comovement of stocks added to and deleted from the S&P 500 Index: Calendar time. Differences between the comovement characteristics of two portfolios of stocks: those about to be added to the S&P 500 and those just recently added. The sample includes stocks added to and deleted from the S&P 500 between 1976 and 2000 which were not involved in mergers or related events (described in the text), and which have sufficient return data on CRSP. A capitalization-weighted return index of non-S&P 500 stocks ($R_{nonSP500}$) in the NYSE, AMEX, and Nasdaq is inferred from the identity described in Table 1. Returns from October 1987 are excluded. In daily data, for example, each day we form an equal-weighted portfolio of stocks that will be added to the S&P 500 within the next year and a portfolio of stocks that were added within in the past year. We then run separate univariate regressions for each portfolio on the S&P 500 index,

$$R_{pre,t} = \alpha_{pre} + \beta_{pre} R_{SP500,t} + \upsilon_{pre,t} \quad \text{and}$$
$$R_{post,t} = \alpha_{post} + \beta_{post} R_{SP500,t} + \upsilon_{post,t},$$

denoting the difference in slope and fit between the "post" and "pre" regressions as $\Delta\beta$ and ΔR^2, respectively. We also run separate bivariate regressions for each portfolio,

$$R_{pre,t} = \alpha_{pre} + \beta_{pre,SP500} R_{SP500,t} + \beta_{pre,nonSP500} R_{nonSP500,t} + \upsilon_{pre,t} \quad \text{and}$$
$$R_{post,t} = \alpha_{post} + \beta_{post,SP500} R_{SP500,t} + \beta_{post,nonSP500} R_{nonSP500,t} + \upsilon_{post,t},$$

denoting the difference in the slopes as $\Delta\beta_{SP500}$ and $\Delta\beta_{nonSP500}$, respectively. The mechanical influence of the pre and post portfolio stocks is removed, as appropriate, from the independent variables. In daily and weekly data, the pre portfolio includes stocks that will be added within one year and the post portfolio includes stocks that were added in the past year. In monthly data, these windows are extended to three years. We require at least 10 stocks in each portfolio in order for that observation (day, week, or month) to be included in the regressions. Asymptotic standard errors are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels in one-sided tests, respectively.

Sample		T	Univariate		Bivariate	
			$\Delta\beta$ (s.e.)	ΔR^2 (s.e.)	$\Delta\beta_{SP500}$ (s.e.)	$\Delta\beta_{nonSP500}$ (s.e.)
Panel A. Daily Returns						
Additions	1976-2000	4147	0.123*** (0.013)	0.100*** (0.007)	0.297*** (0.020)	-0.262*** (0.022)
	1976-1987	1873	0.116*** (0.019)	0. 061*** (0.010)	0.326*** (0.035)	-0.329*** (0.044)
	1988-2000	2274	0.129*** (0.017)	0.119*** (0.010)	0.298*** (0.026)	-0.247*** (0.026)
Deletions	1976-2000	151	0.058 (0.107)	-0.069 (0.063)	0.164 (0.190)	-0.080 (0.117)
Panel B. Weekly Returns						
Additions	1976-2000	856	0.045* (0.026)	0.028* (0.015)	0.167*** (0.046)	-0.154*** (0.043)
	1976-1987	387	0.041 (0.039)	-0.028 (0.024)	-0.014 (0.094)	0.051 (0.099)
	1988-2000	469	0.049 (0.036)	0.065*** (0.022)	0.219*** (0.054)	-0.210*** (0.048)
Deletions	1976-2000	29	-0.082 (0.193)	-0.160 (0.120)	-0.162 (0.537)	0.039 (0.288)
Panel C. Monthly Returns						
Additions	1976-1998	282	0.018 (0.042)	0.090*** (0.030)	0.319*** (0.073)	-0.320*** (0.061)
	1976-1987	127	0.019 (0.047)	0.007 (0.048)	0.148 (0.114)	-0.143 (0.099)
	1988-1998	155	0.016 (0.065)	0.157** (0.067)	0.388*** (0.092)	-0.406*** (0.075)
Deletions	1976-1998	116	-0.123 (0.105)	-0.293*** (0.058)	-0.255* (0.166)	0.132 (0.132)

Table 3. Changes in comovement of stocks added to and deleted from the S&P 500 Index: Relative to matching firms. Changes in the slope and the fit of regressions of returns on stocks added to and deleted from the S&P 500 Index relative to changes in the same parameters of matching stocks. Each stock in the event sample is matched with a stock on industry and growth in market capitalization (described in text) over the pre-change estimation period. The event sample includes stocks added to and deleted from the S&P 500 between 1976 and 2000 which were not involved in mergers or related events (described in the text), which have sufficient return data on CRSP, and for which a matching stock could be found. For each added stock *j*, the univariate model

$$R_{j,t} = \alpha_j + \beta_j R_{SP500,t} + \upsilon_{j,t}$$

and the bivariate model

$$R_{j,t} = \alpha_j + \beta_{j,SP500} R_{SP500,t} + \beta_{j,nonSP500} R_{nonSP500,t} + \upsilon_{j,t}$$

are separately estimated for the pre-change and post-change period, and analogous regressions are run for each matching stock. Returns on the S&P 500 (R_{SP500}) are from the CRSP Index on the S&P 500 Universe file. Returns on a capitalization-weighted index of the non-S&P 500 stocks ($R_{nonSP500}$) in the NYSE, AMEX, and Nasdaq are inferred from the identity described in Table 1. Returns from October 1987 are excluded. The mechanical influence of the added or deleted stock is removed from the independent variables as appropriate. For the univariate regression model, we examine the mean difference between the pre- and post-change slope and fit of the event stock and the matching stock, $\overline{\Delta\Delta\beta}$ and $\overline{\Delta\Delta R^2}$. For the bivariate model, we examine the mean difference between the changes in the slopes of the event stock and the matching stock, $\overline{\Delta\Delta\beta_{SP500}}$ and $\overline{\Delta\Delta\beta_{nonSP500}}$. The pre-change and post-change estimation periods are [-12,-1] and [+1,+12] months for daily and weekly returns and [-36,-1] and [+1,+36] months for monthly returns. Panels A, B, and C show results for daily, weekly, and monthly returns, respectively. Standard errors are determined by simulation, to account for cross-correlation, and are reported in parentheses. ***, **, and * denote statistical significance at the 1%, 5%, and 10% levels in one-sided tests, respectively.

Sample		N	Univariate		Bivariate	
			$\overline{\Delta\Delta\beta}$ (s.e.)	$\overline{\Delta\Delta R^2}$ (s.e.)	$\overline{\Delta\Delta\beta_{SP500}}$ (s.e.)	$\overline{\Delta\Delta\beta_{nonSP500}}$ (s.e.)
Panel A. Daily Returns						
Additions	1976-2000	435	0.120*** (0.022)	0.040*** (0.006)	0.318*** (0.035)	-0.289*** (0.042)
	1976-1987	189	0.109*** (0.028)	0. 033*** (0.008)	0.262*** (0.051)	-0.257*** (0.065)
	1988-2000	246	0.129*** (0.032)	0.046*** (0.007)	0.361*** (0.047)	-0.313*** (0.055)
Deletions	1976-2000	36	-0.098 (0.081)	-0.012 (0.013)	-0.298** (0.142)	0.271 (0.193)
Panel B. Weekly Returns						
Additions	1976-2000	434	0.077** (0.037)	0.028*** (0.009)	0.208*** (0.070)	-0.146** (0.070)
	1976-1987	188	0.086* (0.047)	0.026* (0.015)	0.202* (0.113)	-0.162 (0.114)
	1988-2000	246	0.070 (0.055)	0.030*** (0.012)	0.212** (0.087)	-0.134 (0.088)
Deletions	1976-2000	36	-0.013 (0.157)	-0.030 (0.020)	-0.616** (0.251)	0.771** (0.285)
Panel C. Monthly Returns						
Additions	1976-1998	300	0.104** (0.047)	0.011 (0.013)	0.173* (0.103)	-0.090 (0.091)
	1976-1987	162	0.054 (0.060)	0.008 (0.020)	0.054 (0.145)	-0.011 (0.125)
	1988-1998	138	0.163** (0.073)	0.015 (0.020)	0.313** (0.144)	-0.183 (0.133)
Deletions	1976-1998	18	0.236 (0.156)	0.057 (0.041)	0.438 (0.315)	-0.133 (0.266)

Table 4. Changes in comovement of stocks added to and deleted from the S&P 500 Index: By change in trading volume. Changes in the slope and the fit of regressions of returns of stocks added to and deleted from the S&P 500 Index on returns of the S&P 500 Index and the non-S&P 500 rest of the market. The sample includes stocks added to and deleted from the S&P 500 between 1976 and 2000 which were not involved in mergers or related events (described in the text), and which have sufficient return and trading volume data on CRSP. For each added stock *j*, the univariate model

$$R_{j,t} = \alpha_j + \beta_j R_{SP500,t} + \upsilon_{j,t}$$

and the bivariate model

$$R_{j,t} = \alpha_j + \beta_{j,SP500} R_{SP500,t} + \beta_{j,nonSP500} R_{nonSP500,t} + \upsilon_{j,t}$$

are separately estimated for the pre-change and post-change period. Returns from October 1987 are excluded. The mechanical influence of the added or deleted stock is removed from the independent variables as appropriate. For the univariate regression model, we examine the mean difference between the pre-change slope and the post-change slope $\overline{\Delta\beta}$, and the mean change in fit $\overline{\Delta R^2}$. For the bivariate model, we examine the mean changes in the slopes, $\overline{\Delta\beta_{SP500}}$ and $\overline{\Delta\beta_{nonSP500}}$. The pre-change and post-change estimation samples are [-12,-1] and [+1,+12] months of daily data. Average pre-change and post-change monthly turnover (volume divided by shares outstanding) are computed over these same intervals and used to identify the direction in trading volume. Panel A reports results for stocks that decreased turnover. Panel B reports results for stocks that increased turnover. Standard errors are determined by simulation, to account for cross-correlation, and are reported in parentheses. ***, **, and * denote significant differences from zero at the 1%, 5%, and 10% levels in one-sided tests, respectively.

Sample		N	Univariate		Bivariate	
			$\overline{\Delta\beta}$ (s.e.)	$\overline{\Delta R^2}$ (s.e.)	$\overline{\Delta\beta_{SP500}}$ (s.e.)	$\overline{\Delta\beta_{nonSP500}}$ (s.e.)
Panel A. Turnover decrease						
Additions	1976-2000	151	0.083** (0.037)	0.062*** (0.009)	0.313*** (0.048)	-0.374*** (0.055)
	1976-1987	65	0.054* (0.036)	0.052*** (0.013)	0.248*** (0.073)	-0.294*** (0.086)
	1988-2000	86	0.105** (0.060)	0.070*** (0.012)	0.362*** (0.063)	-0.434*** (0.071)
Panel B. Turnover increase						
Additions	1976-2000	295	0.194*** (0.025)	0.045*** (0.007)	0.350*** (0.033)	-0.307*** (0.042)
	1976-1987	122	0.089*** (0.031)	0.035*** (0.010)	0.344*** (0.050)	-0.342*** (0.062)
	1988-2000	173	0.269*** (0.036)	0.052*** (0.009)	0.354*** (0.044)	-0.282*** (0.056)

Table 5. Changes in comovement of stocks added to and deleted from the S&P 500 Index: Information diffusion effects (5 leads and lags). Changes in the slope and the fit of regressions of daily returns on stocks added to and deleted from the S&P 500 Index on daily returns of the S&P 500 Index and the non-S&P 500 rest of the market, using five leads and lags in daily returns to adjust beta for information diffusion effects as suggested by Dimson (1979). The sample includes stocks added to and deleted from the S&P 500 between 1976 and 2000 which were not involved in mergers or related events (described in the text), and which have sufficient return data on CRSP. For each added stock *j*, the univariate model

$$R_{j,t} = \alpha_j + \sum_{s=-5}^{5} \beta_j^{(s)} R_{SP500,t+s} + \upsilon_{j,t}$$

and the bivariate model

$$R_{j,t} = \alpha_j + \sum_{s=-5}^{5} \left(\beta_{j,sp500}^{(s)} R_{SP500,t+s} + \beta_{j,nonSP500}^{(s)} R_{nonSP500,t+s} \right) + \upsilon_{j,t}$$

are separately estimated for the pre-change and post-change period. Returns on the S&P 500 (R_{SP500}) are from the CRSP Index on the S&P 500 Universe file. Returns on a capitalization-weighted index of the non-S&P 500 stocks ($R_{nonSP500}$) in the NYSE, AMEX, and Nasdaq are inferred from the identity described in Table 1. Returns from October 1987 are excluded. The mechanical influence of the added or deleted stock is removed from the independent variables as appropriate. In Panel A, we report the mean difference between the pre-change and post-change Dimson beta (which is defined as the sum of the lag, contemporaneous, and lead beta coefficients). In Panel B, we report the mean difference between the pre-change and post-change components of the Dimson beta. Standard errors are determined by simulation, to account for cross-correlation, and are reported in parentheses. ***, **, and * denote significant differences from zero at the 1%, 5%, and 10% levels in one-sided tests, respectively.

Sample		N (Panel A) / term (Panel B)	Univariate $\overline{\Delta\beta}$ (s.e.)	Univariate $\overline{\Delta R^2}$ (s.e.)	Bivariate $\overline{\Delta\beta_{SP500}}$ (s.e.)	Bivariate $\overline{\Delta\beta_{nonSP500}}$ (s.e.)
Panel A. Dimson Beta (5 leads and lags)						
Additions	1976-2000	449	0.109***	0.038***	0.104	-0.007
			(0.046)	(0.005)	(0.090)	(0.082)
	1976-1987	190	0.018	0.026***	0.036	0.002
			(0.053)	(0.008)	(0.156)	(0.142)
	1988-2000	259	0.175***	0.047***	0.154*	-0.013
			(0.066)	(0.007)	(0.106)	(0.097)
Deletions	1976-2000	76	-0.020	-0.012*	-0.631**	0.569***
			(0.162)	(0.008)	(0.342)	(0.278)

Panel B. Components of Dimson Beta (5 leads and lags)					
Additions	1976-2000	*t-5*	-0.002	-0.025	0.063***
			(0.012)	(0.021)	(0.027)
		t-4	0.035***	0.054***	-0.041*
			(0.012)	(0.023)	(0.030)
		t-3	0.009	-0.002	0.019
			(0.012)	(0.024)	(0.029)
		t-2	-0.005	-0.042**	0.024
			(0.024)	(0.023)	(0.028)
		t-1	-0.080***	-0.118***	0.151***
			(0.014)	(0.025)	(0.031)
		t	0.150***	0.310***	-0.310***
			(0.021)	(0.028)	(0.037)
		t+1	-0.006	-0.071***	0.069**
			(0.012)	(0.025)	(0.032)
		t+2	0.007	-0.055**	0.058**
			(0.010)	(0.024)	(0.029)
		t+3	-0.020**	-0.034*	0.051**
			(0.011)	(0.022)	(0.027)
		t+4	0.021**	0.044**	-0.026
			(0.011)	(0.022)	(0.022)
		t+5	0.000	0.042**	-0.065**
			(0.012)	(0.022)	(0.029)

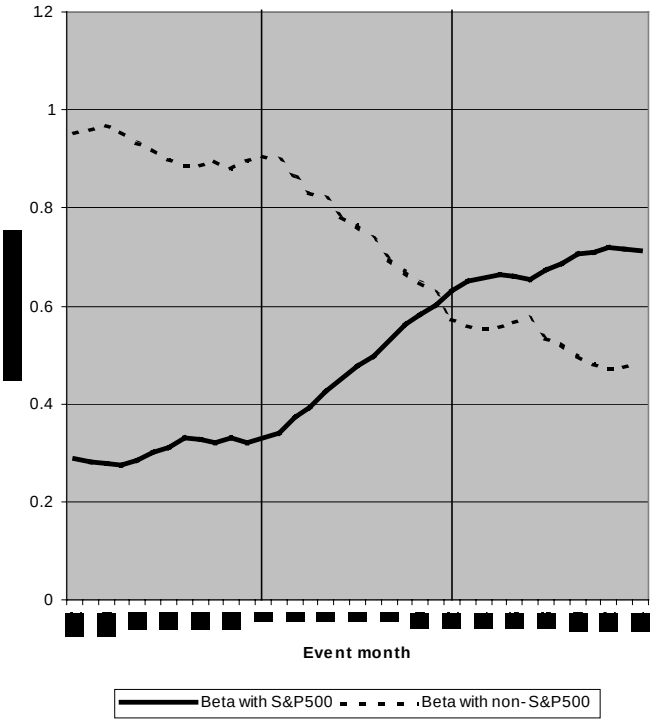
Figure I. Changes in comovement of stocks added to the S&P 500 Index and stocks with matching characteristics. Plots of the mean slope coefficients of bivariate regressions of returns on stocks added to the S&P 500, and stocks with matching characteristics to those added, on returns of the S&P 500 Index and the non-S&P 500 rest of the market. The event sample includes stocks added to the S&P 500 which were not involved in mergers or related events (described in the text), which have complete returns data over the entire event horizon examined in each figure (-12 to +24 months in daily and weekly returns data and –36 to +72 months in monthly returns data), which remain in the Index for the full post-event horizon, and for which suitable matching firms exist that have complete data over the same horizon. Each stock in the event sample is matched with a stock on industry and growth in market capitalization (as described in text) over the pre-change estimation period. For each added stock *j* (and each corresponding matching stock), the bivariate model

$$R_{j,t} = \alpha_j + \beta_{j,SP500} R_{SP500,t} + \beta_{j,nonSP500} R_{nonSP500,t} + \upsilon_{j,t}$$

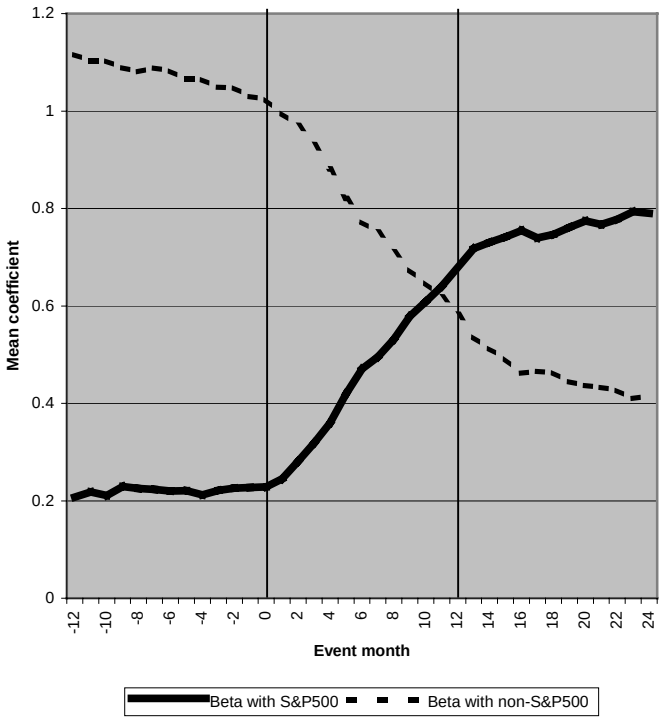
is estimated in rolling regressions where sample intervals are [-12,-1] months for daily and weekly returns regressions and [-36,-1] months for monthly returns. Returns on the S&P 500 (R_{SP500}) are from the CRSP Index on the S&P 500 Universe file. Returns on a capitalization-weighted index of the non-S&P 500 stocks ($R_{nonSP500}$) in the NYSE, AMEX, and Nasdaq are inferred from the identity described in Table 1. Returns from October 1987 are excluded. The mechanical influence of the added stock is removed, as appropriate, from both independent variables. The means of the event stock coefficients are plotted in event time in the top half of each panel, and the means of the matching stock coefficients are plotted in the bottom half. The left vertical line indicates the addition date; coefficients to the left of this line are estimated using only pre-event data. Coefficients to the right of the right vertical line are estimated using only post-event data. In between, coefficients are estimated using both pre- and post-event data. Panels A, B, and C show results for daily, weekly, and monthly returns, respectively.

A. Daily Returns

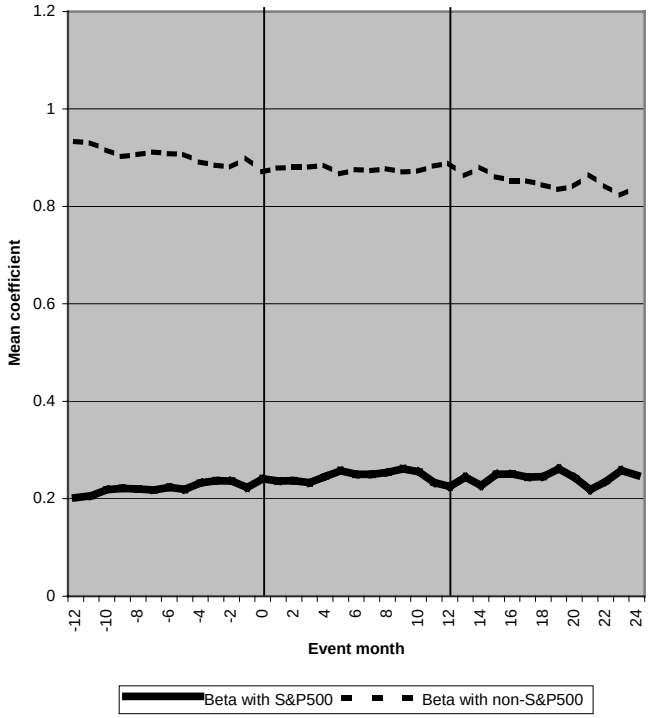
1976-1987 additions with matches (N = 169)



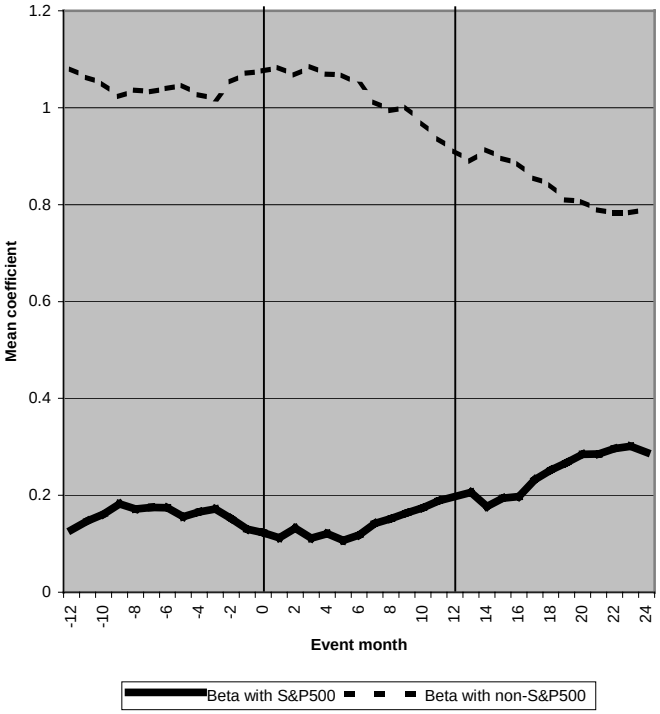
1988-2000 additions with matches (N = 153)



1976-1987 matching firms (N = 169)

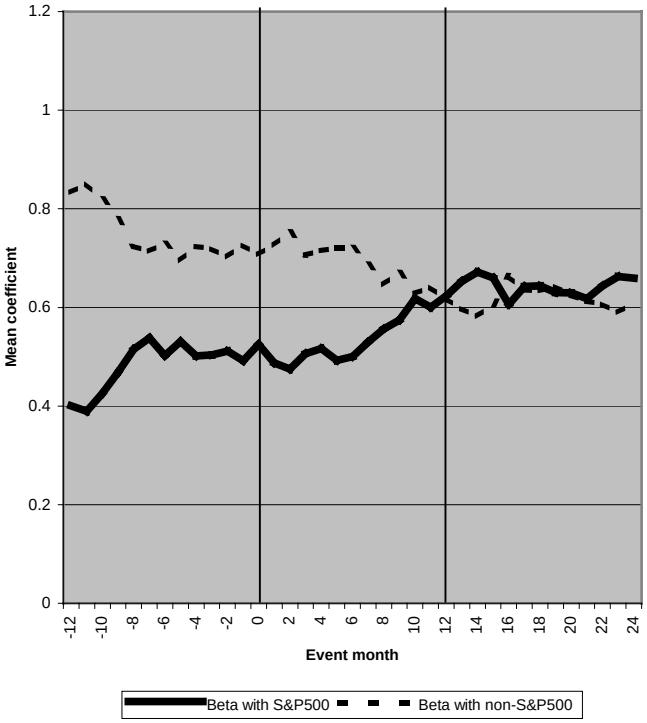


1988-2000 matching firms (N = 153)

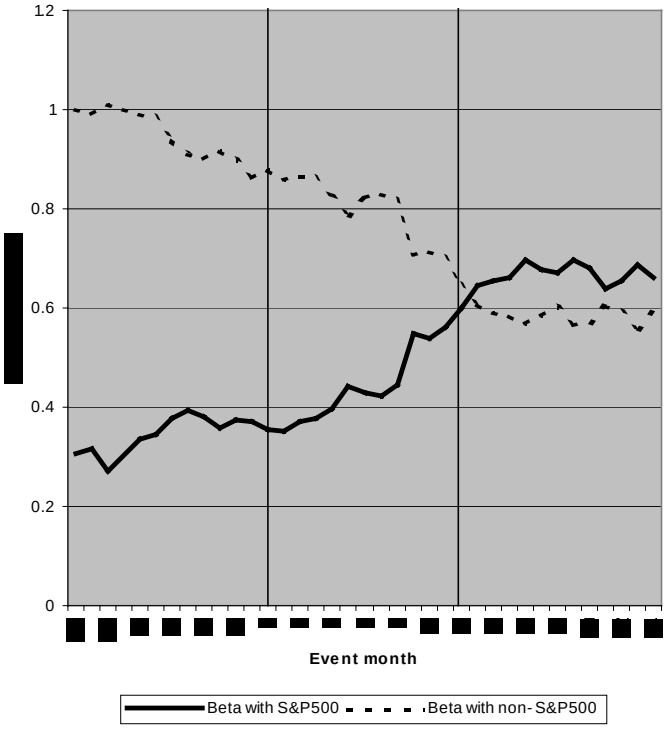


B. Weekly Returns

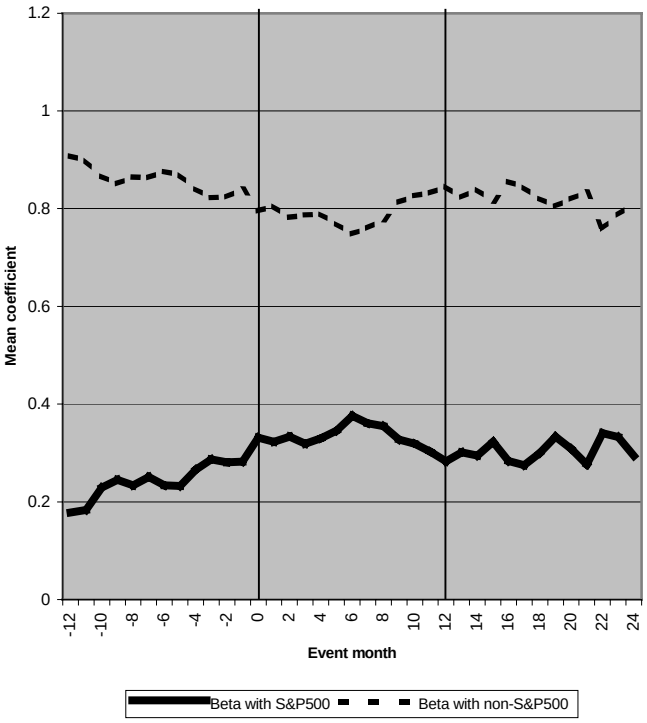
1976-1987 additions with matches (N = 169)



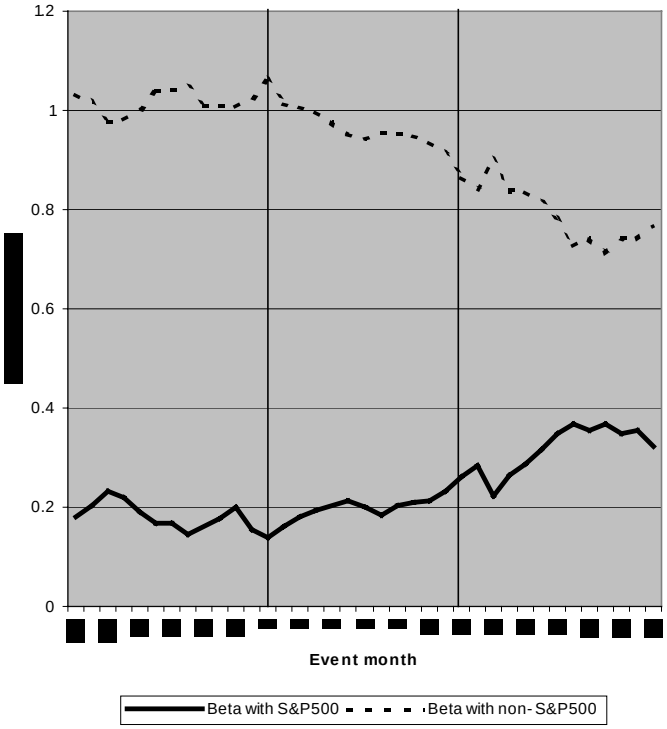
1988-2000 additions with matches (N = 153)



1976-1987 matching firms (N = 169)

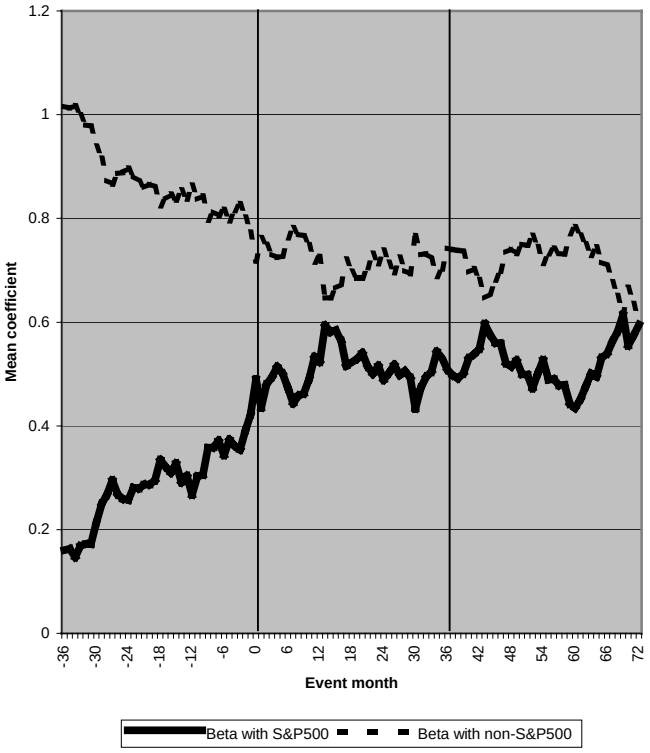


1988-2000 matching firms (N = 153)

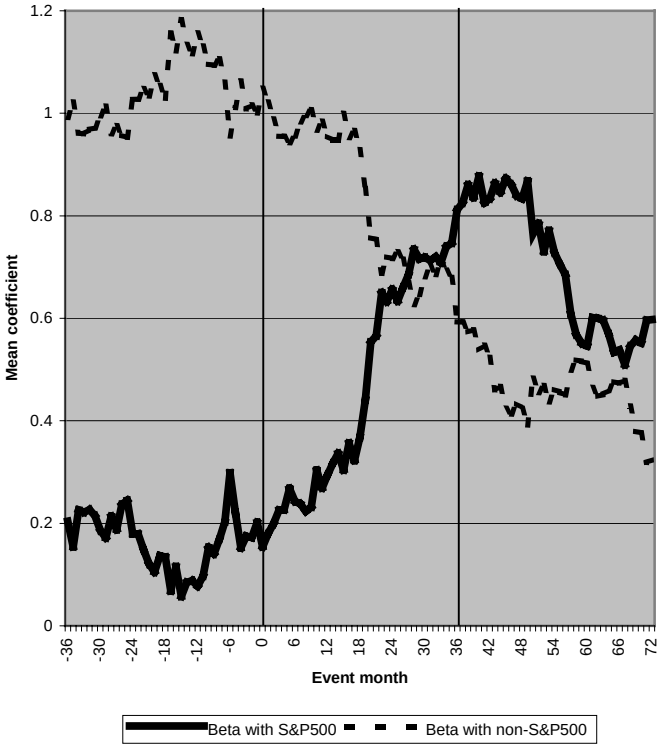


C. Monthly Returns

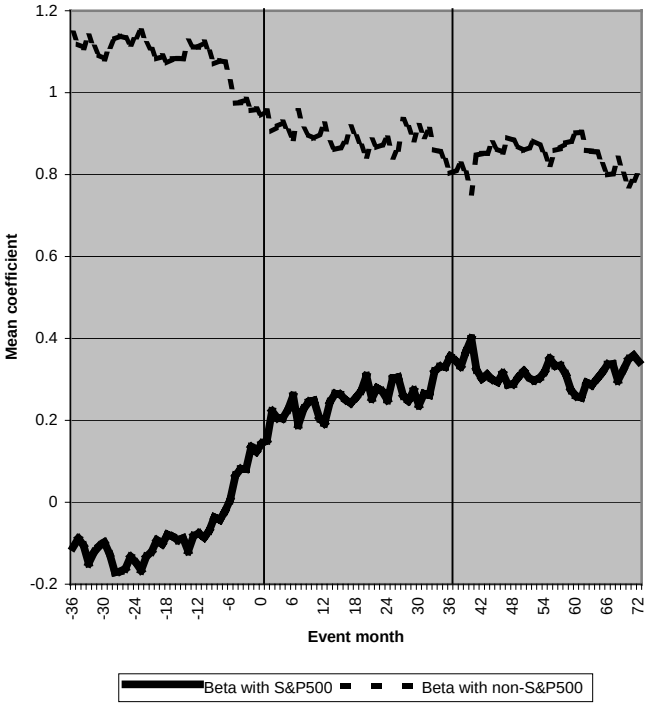
1976-1987 additions with matches (N = 97)



1988-2000 additions with matches (N = 47)



1976-1987 matching firms (N = 97)



1988-2000 matching firms (N = 47)