Cash-Flow Driven Covariation

Miguel Antón*
IESE Business School, Universidad de Navarra
manton@iese.edu

First draft: November, 2010
This draft: January, 2013

Abstract

This paper studies the sources of change in the systematic risks of stocks added to the S&P 500 index. Firstly, using vector autoregressions (VARs) and a two-beta decomposition, I measure the different components of beta before and after the addition. I find that I cannot reject the hypothesis that all of the well-known change in beta comes from the cash-flow news component of a firm's return. Secondly, I study fundamentals of included firms directly to reduce any concerns that the VAR-based results are sensitive to my particular specification. This analysis confirms that post inclusion, the profitability of a company added to the index varies significantly more with the profitability of the S&P 500. As ownership structure cannot directly influence fundamentals, these results challenge previous findings, as they are consistent with the change in beta being due to a selection effect.

*IESE Business School, Universidad de Navarra, Av. Pearson 21, 08034 Barcelona, Spain. I am extremely grateful to Christopher Polk for his constant support and guidance. I also thank Nick Barberis, Dan Bergstresser, Cecilia Bustamante, Mikhail Chernov, Vicente Cuñat, Vincent Fardeau, Daniel Ferreira, Fabian Garavito, Christian Julliard, Dong Lou, Dimitri Vayanos, Michela Verardo, David Webb, and Kathy Yuan for helpful comments and suggestions.

1 Introduction

In standard finance models fundamentals drive asset prices. There is however a large body of the literature documenting departures of prices from fundamentals[1]. It is difficult to explain under the traditional paradigm market anomalies (e.g. momentum, reversal, value effect). Some of the evidence interpreted as favouring non-fundamental-based theories concerns index effects, both in first and second moments. For instance, Vijh (1994) and Barberis, Shleifer and Wurgler (2005) find that index additions are followed by an increase in covariation, and argue that this effect is not driven by fundamentals.

Index additions have been widely used as a quasi-natural experiment to distinguish between competing theories. For example, a number of papers show that there is a significant jump in price levels following index additions and deletions[2]. Much of these findings have been interpreted as evidence of non-fundamental-based theories. Some research, however, have challenged the interpretation of this effect. Dennis et al. (2003) for example argue that index additions are not fully information-free events, as they are followed by increases in earnings. While the interpretation of these effects in the first moments has been subject to debate among academics, changes in second moments (covariances) around index inclusions are widely accepted as evidence of non-fundamental-based theories[3].

In this paper I show that S&P 500 index inclusions are followed by changes in cash-flow covariances. I specifically take on the task of disentangling how much of the change in beta after an index addition corresponds to a fundamental effect and how much to a non-fundamental effect. I provide evidence of changes in cash-flow news' covariances after index additions using a two beta decomposition. Following Campbell and Mei (1993), I decompose beta into discount-rate and cash-flow shocks of the individual firm with the market. I find that I cannot reject the hypothesis that all of the well-known change in beta comes from the cash-flow news component of a firm's return. As investors cannot directly influence fundamentals, these results challenge previous findings, as they are consistent with the change in beta being due

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

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

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

to a selection effect.

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

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

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

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

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

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

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

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

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

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

2 Decomposing Stock Returns and Betas

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

2.1 Decomposing Returns

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

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

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

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

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

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

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

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

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

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

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

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

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

2.2 Decomposing Betas

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

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

and

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

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

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

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

3 A VAR framework

3.1 Measuring the components of returns

I use vector autoregressions (VARs) to measure the shocks to cash flows and to discount rates, following Campbell (1991) approach. The VAR methodology first estimates the terms $\mathrm{E}_t \, r_{t+1}$ and $(\mathrm{E}_{t+1} - \mathrm{E}_t) \sum_{j=1}^{\infty} \rho^j r_{t+1+j}$ and then uses realization of r_{t+1} and equation 2.4 to back out cash-flow news. Because of the approximate identity linking returns, dividends, and stock prices, this approach yields results that are almost identical to those that are obtained by forecasting cash flows explicitly using the same information set. Thus the choice of variables to enter the VAR is the important decision in implementing this methodology.

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

$$z_{t+1} = a + \Gamma z_t + u_{t+1}, \tag{3.1}$$

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

Assuming that the process in equation (3.1) generates the data, $t+1$ cash-flow and discount-rate news are linear functions of the $t+1$ shock vector:

$$\begin{aligned} N_{DR,t+1} &= e1'\lambda u_{t+1}, \\ N_{CF,t+1} &= (e1' + e1'\lambda)\, u_{t+1}. \end{aligned} \tag{3.2}$$

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

3.2 Aggregate VAR Specifications

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

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

The first element the VAR is the excess return on the market (r_m^e), calculated as the difference between the monthly log return on the CRSP value-weighted stock index (r_m) and the monthly log risk-free rate (r_f). I take the excess return series from Kenneth French's website[4]. The second element in the VAR is the term yield spread (TY), provided by Global Financial Data and computed as the yield difference between ten-year constant-maturity taxable bonds and short-term taxable notes, in percentage points[5]. The third variable is the log smoothed price-earnings ratio (PE), the log of the price of the S&P 500 index divided by a ten-year trailing moving average of aggregate earnings of companies in the index. I take the price-earnings ratio series from Robert Shiller's website[6]. As in Campbell and Vuolteenaho (2004), I carefully remove the interpolation inherent in Shiller's construction of the variable to ensure the variable does not suffer from look-ahead bias. The final variable is the small-stock value spread (VS), which I construct using the data made available by Professor Kenneth French on his web site. The portfolios, which are constructed at the end of each June, are the intersections of two portfolios formed on size (market equity, ME) and three portfolios formed on the ratio of book equity to market equity (BE/ME). I generate intermediate values of VS by accumulating total returns on the portfolios in question.

The motivation for the use of these variables is the following. Term yield spread tracks the business cycle, as pointed out by Fama and French (1989), and there are several reasons why we should expect aggregate returns to be correlated to the business cycle. Second, if price-earnings ratio is high and expected earnings growth is constant, then long-run expected returns must be low, so we expect a

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

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

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

negative coefficient of this variable in the VAR. Finally, the small-stock value spread is included given the evidence in Brennan, Wang, and Xia (2001) and others that relatively high returns for small growth stocks predict low aggregate returns in the market.

Table 1 reports the VAR model parameters for the aggregate VAR, estimated using OLS. Every row of the table corresponds to a different equation of the VAR. The first five columns report coefficients on the five explanatory variables: a constant, and lags of the excess market return, term yield spread, price-earnings ratio, and small-stock value spread. OLS standard errors are reported in parentheses below the coefficients.

The first row in Table 1 shows that all four of my VAR state variables have some ability to predict monthly excess returns on the market excess returns. Monthly market returns display momentum; the coefficient on the lagged market excess return is a statistically significant 0.1118 with a t-statistic of 3.52.

The regression coefficient on past values of the term yield spread is positive, consistent with the findings of Keim and Stambaugh (1986), Campbell (1987), and Fama and French (1989), but with a t-statistic of 1.6. As expected, the smoothed price-earnings ratio negatively predicts market excess returns, with t-statistics of 3.41, consistent with the finding that various scaled-price variables forecast aggregate returns (Campbell and Shiller, 1988ab, 2003; Rozeff 1984; Fama and French 1988, 1989). Finally, the small-stock value spread negatively predicts market excess returns with t-statistics of 2.16, consistent with Brennan, Wang, and Xia (2001), Eleswarapu and Reinganum (2004), and Campbell and Vuolteenaho (2004). The estimated coefficients, both in terms of signs and t-statistics, are consistent with previous research.

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

3.3 Firm-level VAR Specification

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

I measure BE for the fiscal year ending in calendar year $t-1$, and ME (market value of equity) at the end of May of year t[7]. I update BE/ME over the subsequent eleven months by dividing by the cumulative gross return from the end of May to the month in question. I require each firm-year observation to have a valid past BE/ME ratio that must be positive in value. Moreover, in order to eliminate likely data errors, I censor the BE/ME variables of these firms to the range (.01,100) by adjusting the book value. To avoid influential observations created by the log transform, I first shrink the BE/ME towards one by defining $BM \equiv \log[(.9BE + .1ME)/ME]$.

The firm-level VAR generates market-adjusted cash-flow and discount-rate news for each firm and month. I remove month-specific means from the state variables by subtracting $r_{M,t}$ from $r_{i,t}$ and cross-sectional means from $MOM_{i,t}$ and $BM_{i,t}$. As in Campbell, Polk, and Vuolteenaho (2010), instead of subtracting the equal-weight

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

cross-sectional mean from $r_{i,t}$, I subtract the log value-weight CRSP index return, because this will allow us to undo the market adjustment simply by adding back the cash-flow and discount-rate news extracted from the aggregate VAR.

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

Parameter estimates, presented in Table 2, imply that expected returns are high when past one-month return is low and when the book-to-market ratio and momentum are high. Book-to-market is the statistically most significant predictor, while the firm's own stock return is the statistically least significant predictor. Momentum is high when past stock return and past momentum are high and the book-to-market ratio is low. The book-to-market ratio is quite persistent. Controlling for past book-to-market, expected future book-to-market ratio is high when the past monthly return is high and past momentum is low.

4 Empirical Results

4.1 Data

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

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

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

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

4.2 Changes in Betas in a VAR Framework

4.2.1 Benchmark case

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

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

$$r^e_{i,t} = \alpha_i + \beta^b_{SP} r^e_{SP,t} + \beta^b_{nSP} r^e_{nSP,t} + \Delta\beta_{SP} I_{it} r^e_{SP,t} + \Delta\beta_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \qquad (4.1)$$

The coefficients of the interactions $I_{it} * r^e_{SP,t}$ and $I_{it} * r^e_{nSP,t}$ ($\Delta\beta_{SP}$ and $\Delta\beta_{nSP}$

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

$$r_{M,t} = \left(\frac{CAP_{M,t-1} - CAP_{SP,t-1}}{CAP_{M,t-1}}\right) r_{nSP,t} + \left(\frac{CAP_{SP,t-1}}{CAP_{M,t-1}}\right) r_{SP,t} \tag{4.2}$$

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

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

Table 3 shows the results for this regression. Consistent with previous literature (Barberis, Shleifer, and Wurgler, 2005), I find that beta with respect to S&P 500 returns jumps and beta with respect to non S&P 500 returns falls, both significantly. The second row displays the average change in S&P 500 beta, $\Delta\beta_{SP}$, 0.425, accurately estimated with a t-stat of 6.25. The fourth row shows the average change in non S&P 500 beta, $\Delta\beta_{nSP}$, with the coefficient -0.291, estimated with a a t-stat of 4.59.

4.2.2 Cash-flow and discount-rate betas

The results reported in Table 3, in line with those found by Barberis et. al, have been interpreted as evidence of friction- or sentiment-based comovement. The argument is the following. Standard and Poors state clearly that in choosing a company to be included in the index, they do not signal anything about the future performance of the company. As a consequence, any change in the betas of companies added to the index should be attributed to sentiment, because fundamentals have not changed.

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

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

To implement this test, I simply substitute the excess returns of event stocks, $r^e_{i,t}$, for their cash-flow news ($N_{iCF,t}$) and (negative of) discount-rate news ($-N_{iDR,t}$) in the left-hand side of equation 4.1:

$$-N_{iDR,t} = \alpha_i + \beta^{DRb}_{SP} r^e_{SP,t} + \beta^{DRb}_{nSP} r^e_{nSP,t} + \Delta\beta^{DR}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{DR}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \quad (4.3)$$

and

$$N_{iCF,t} = \alpha_i + \beta^{CFb}_{SP} r^e_{SP,t} + \beta^{CFb}_{nSP} r^e_{nSP,t} + \Delta\beta^{CF}_{SP} I_{it} r^e_{SP,t} + \Delta\beta^{CF}_{nSP} I_{it} r^e_{nSP,t} + \varepsilon_{i,t} \quad (4.4)$$

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

$$\begin{aligned} \Delta\beta_{SP} &\approx \Delta\beta^{DR}_{SP} + \Delta\beta^{CF}_{SP} \\ \Delta\beta_{nSP} &\approx \Delta\beta^{DR}_{nSP} + \Delta\beta^{CF}_{nSP} \end{aligned} \quad (4.5)$$

Table 4 shows the changes in cash-flow and discount rate betas. The first column replicates the benchmark column of table 3. The second and third columns show the results for the change in the different beta components. The change in discount rate beta with respect to the S&P 500 is an insignificant -0.008 (second row, second column), and 0.049 with respect to the non S&P 500 stocks, whereas the changes in cash-flow betas are 0.391 and -0.286 (for S&P 500 and non S&P 500 respectively), accurately estimated with t-stats of 6.15 and 4.62. This result strongly supports the idea that, at the monthly frequency, sentiment- or friction-based comovement is negligible if not inexistent.

Figure 1 shows the evolution of average betas around the inclusion event. Rolling regressions are estimated with windows of 36 months from month -36 to month $+72$. In the top panel we observe the evolution of the overall average betas. S&P 500 betas increase significantly after inclusion, and non S&P 500 decrease after inclusion. Below, in the central panel, rolling average discount rate betas are plotted, showing a very mild pattern of variation. Finally, in the bottom panel, we see how all the action in the change in beta is originated in the cash-flow betas.

4.3 Results from a direct approach

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

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

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

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

I run a pooled-OLS quarterly regression. Results are presented in table 5. The results confirm my findings in the VAR approach. When a stock is not in the index, its beta with S&P 500 return on equity is 0.227 and its beta with the rest of the market return on equity is 0.716, with both coefficients estimated precisely with t-statistic above 3. However, once the stock has been added to the index, the betas turn to 0.488 and 0.211 for S&P 500 and rest of the market return on equities.

4.4 Matched stocks

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

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

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

Table 6 shows the results of the changes in beta using matched stocks. I find that matched stocks exhibit similar patterns in betas, as matched stocks also experience a significant change in beta with respect to S&P 500 returns, of 0.261. The crucial result in this table is that the difference in difference in betas, though mildly significant (0.165 with a t-stat of 1.91), it all comes from the cash-flow component: 0.158 with a t-stat of 2. This is both evidence of Standard and Poors signaling something about future cash-flow covariances, and of Standard and Poors' Committee being a better predictor of future cash-flow covariances and relevance in the economy than the basic and always imperfect matching algorithm that we employ.

Figure 2 shows the evolution of rolling average betas (for the overall betas, and their discount-rate and cash-flow components). The top panel shows the betas for the event firms (those included in the S&P 500), and the bottom panel shows the evolution of betas for matched firms (firms that could have been included in the index, but were not).

4.5 Reconciling with Barberis et al.

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

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

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

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

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

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

4.6 Robustness to different subsamples

4.6.1 Subsample in the time dimension

I explore the effect in different time subsamples to uncover effects that might be hidden in the full-sample period. Previous research has found that the change in beta after index additions has grown over time. Consistent with those findings, I find that the effect is stronger in the last part of the sample. This analysis, shown in table 7, reflects three findings. Firstly, the effect of the change in beta with respect to S&P 500 index comes from the cash-flow components of the stocks added rather from the discount rates in both parts of the subsample. The changes in beta for the two subsamples are 0.230 and 0.533, estimated with t-stats above 3, where almost all the effect is cash-flow originated (0.297 and 0.393).

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

4.6.2 Subsample in growth value dimension

In this subsection I study whether stocks with different characteristics differ in the change in beta experienced after inclusion. I divide the included firms into growth and value stocks, by comparing the cross-sectionally adjusted book-to-market ratios. Growth firms tend to be more intangible and more opaque, while value firms are more stable, if they are financially sound. Because the change in beta also reflects the size of the companies added, growth stocks should exhibit a higher change in beta than value stocks. Table 8 reports the results. Consistent with my prior, I find that the change in beta is higher for growth firms (0.547 versus 0.356). The results for matched firms exhibit similar patterns, and the difference in difference, although insignificant, is also coming from the cash-flow components of beta.

4.7 Robustness to a second-order VAR

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

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

$$z_{t+1} = a + A_1 z_t + A_2 z_{t-1} + u_{t+1} \tag{4.6}$$

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

$$\begin{bmatrix} z_{t+1} \\ z_t \end{bmatrix} = \begin{bmatrix} a \\ 0 \end{bmatrix} + \begin{bmatrix} A_1 & A_2 \\ I & 0 \end{bmatrix} \begin{bmatrix} z_t \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} u_{t+1} \\ 0 \end{bmatrix} \tag{4.7}$$

Table 9 shows the results for the second-order aggregate VAR. To avoid an unncessary display of zeros and the identity matrix, I only show A_1 and A_2. The results are similar to the first-order VAR. Due to the additional free parameters, however, the standard errors are somewhat larger. The coefficients for the second lag are estimated less accurately. Market returns exhibit now a bit of reversal in the second lag (with a coefficient of -0.04), term yield spread and price earning ratio keeps the positive sign in the second lag estimate, and the small stock value spread flips sign with respect to the first lag. The intercepts and the R-Squares are very similar to the previous specification.

I now turn to the firm-level market adjusted VAR. The variables used in the benchmark first-order VAR are the following: market adjusted log stock return (r_i), the previous year return, excluding the last month (MOM_i), and the log book-to-market (BM_i). I motivate this lag order as a second-order cointegrating VAR. Previous research has also shown that these variables have predictive power beyond the first month (Vuolteenaho, 2002, and Campbell, Polk, and Vuolteenaho, 2010). Consistent with Vuolteenaho (2002), I find that the results are very similar to the first-order VAR. Table 10 shows the coefficients of the secon-order market-adjusted firm-level VAR. As in the aggregate VAR, the standard errors of the second-order coefficients are large, and thus the coefficients are not accurately estimated. Monthly returns also exhibit reversal in the second lag, and the previous year return computed in the second lag predicts also positively the returns. The coefficient for book-to-market shows a different sign for the second lag, which is consistent with the first-order VAR given the degree of correlation between the book-to-market at time t and the book-to-market at time $t-1$.

Following the same methodology, I extract the news from each of the new VARs (the second-order aggregate VAR and second-order firm-level VAR), I add them up, yielding $N_{iDR,t}$ and $N_{iCF,t}$, and compute the changes in cash flow and discount rate betas after the addition in the S&P 500 index, as before. Table 11 shows the changes in overall beta (which I include again for comparison purposes), and the changes in the new cash flow and new discount rate betas. The main results are very robust to the use of a second order VAR. In column three we observe that the change in beta after an S&P 500 addition comes from the cash-flow beta. The overall change in beta is a strongly significant 0.430, the change in discount rate beta is an insignificant -0.035, and the change in cash flow beta is a stronly significant 0.424. Consistent with the results from the first order VAR, matched stocks also experience a change in the cash flow betas (column 6), and the difference in difference is all coming from the cash-flows (see column 9), although it is estimated less accurately.

4.8 An alternative specification of the VAR

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

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

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

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

Several specifications of the model are possible. In Table 12, I show the different options. This table only shows the first equation of the VAR for the different specific-ations (where the dependent variable is the firm-level excess log return). Firm-level variables include the excess log return, the previous year return (excluding the last month) in excess of the risk free rate during the same period, the log book to market ratio, and the log profitability in excess of the risk free rate. I include two sets of market-wide variables. The first one comprises the cross-sectional medians of the firm-level state variables, and the second one includes the four aggregate variables used to estimate the aggregate VAR in the previous section.

Columns (1) and (2) in Table 12 show the results when including the two different blocks of market-wide variables. In column (1) we can observe that all market-wide variables have predictive power consistent with previous literature, except the cross-sectional median of the variable MOM_t. In column (2) we also observe that all the aggregate variables have some predictive power as well, though not all them very significant. In order to have a relatively parsimonious VAR and choose the most significant variables, I conduct a horse-race of all the variables, as shown in column (3). Once all eight market-wide variables are included, we can see that three of the four cross-sectional medians cease to be significant, whereas the market return and term yield spread still have explanatory power. Although the cross-sectional median of profitability is significant in this specification, it appears insignificant if the insignificant variables are dropped (this and other horse-race options have been evaluated but not shown for the sake of brevity). The final set of variables I use are the ones shown in column (4).

Table 13 shows all the coefficients for the VAR corresponding to column (4) in the previous table. Intercepts are included in the VAR, however the magnitude is very small and insignificant in all cases. All state variables in the first equation are significant at the 1%. The sign of the variables is as expected: the coefficient for excess log return is negative (showing the short-term reversal), positive and strong for momentum, profitability, market return and term yield spread. The equations corresponding to the aggregate variables are consistent with the aggregate VAR estimated in the previous section: market return exhibits momentum at the monthly level, and term yield spread predicts positively market return. The R-Square, 2%, is also similar (although lower, because there are only two variables predicting market returns now) to the previous aggregate VAR 2.81%.

I then extract the news from this new VAR, $N_{iDR,t}$ and $N_{iCF,t}$, and compute the changes in cash flow and discount rate betas after the addition in the S&P 500 index, as before. The only difference is that I now estimate the betas with different news, the ones extracted from this alternative specification of the VAR. Table 14 shows the changes in overall beta (included again for comparison purposes), in the new cash-flow and discount rate betas. The main results are robust to this different specification of the VAR. In column three we observe that the change in beta after an S&P 500 addition comes from the cash-flow beta. The overall change in beta is a stronly significant 0.430, the change in discount rate beta is an insignificant 0.036, and the change in cash flow beta is a stronly significant 0.357. And as in the previous Section, when compared the changes in betas with matched stocks, the difference in difference is all coming from the cash-flows, and is less significant than for the event stocks.

4.9 Alternative cash flow risk measure

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

5 Conclusion

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

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

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

References

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

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

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

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

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

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

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

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

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

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

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

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

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

Cohen, R., Christopher Polk, and Tuomo Vuolteenaho, 2003, The Value Spread, *Journal of Finance* 58, 609-641.

Cohen, R., Christopher Polk, and Tuomo Vuolteenaho, 2010, The Prices is (almost) Right, *Journal of Finance* 64, 2739-2782.

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

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

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

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

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

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

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

Gromb, Denis and Dimitri Vayanos, 2010, Limits of Arbitrage: The State of the Theory, *Annual Review of Financial Economics* 2, 251-275.

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.

Keim, D., and R. Stambaugh, 1986, Predicting Returns in the Stock and Bond Markets, *Journal of Financial Economics* 17, 357-390.

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

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

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

Vijh, A., 1994, S&P 500 trading strategies and stock betas, *Review of Financial Studies* 7, 215-251.

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

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

Table 1: **Aggregate VAR**

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

Aggregate VAR to predict market return

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

Table 2: **Firm-level VAR**

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

Firm-level VAR for market-adjusted returns

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

Table 3: **Changes in Beta - Benchmark Case**

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

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

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

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

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

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

Table 4: **Changes in cash-flow and discount rate betas**

This table shows the changes in the slope of regressions of returns (and its components) of stocks added to the S&P 500 on returns of the S&P 500 index and the non-S&P 500 rest of the market. The sample and definition of variables is described in Table 3. This table shows the results of regressions similar to the previous table, but replacing the returns on the left hand side variable with (negative of) discount-rate news ($-N_{i,DR}$) and cash-flow news ($N_{i,CF}$) of the event stocks. The equations estimated are the following:

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

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

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

Table 5: **Direct measures of cash flows**

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

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

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

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

Table 6: **Changes in beta and matching firms**

This table shows the changes in the slope of regressions of returns (and its components) of stocks added to the S&P 500, matched stocks, and their difference, on returns of the S&P 500 index and the non-S&P 500 rest of the market. Firms are matched to event stocks based on industry and size, as described in the text. The sample and definition of variables is described in Table 3. The equations estimated are the following:

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

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

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

Table 7: **Robustness to time subsamples**

This table follows the same sample and procedure as table 6. Panel A shows the results using only the bottom quintile of event stocks sorted on book-to-market ratios. Panel B shows the results using only the top quintile of event stocks sorted on book-to-market ratios. The equations estimated are similar to those in table 6. I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

1976-1991	Event Firms			Matched Firms			Difference		
	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.375***	0.793***	-0.420***	0.488***	0.772***	-0.297***	-0.106	0.022	-0.119*
	(0.065)	(0.072)	(0.085)	(0.062)	(0.067)	(0.080)	(0.071)	(0.023)	(0.067)
$I_{it}r^e_{SP,t}$	0.230***	-0.077*	0.297***	-0.043	-0.061	0.038	0.266***	-0.016	0.253***
	(0.067)	(0.045)	(0.066)	(0.071)	(0.043)	(0.069)	(0.096)	(0.026)	(0.080)
$r^e_{nSP,t}$	0.733***	0.106*	0.501***	0.578***	0.125**	0.354***	0.150**	-0.021	0.143**
	(0.054)	(0.061)	(0.076)	(0.054)	(0.057)	(0.077)	(0.060)	(0.020)	(0.055)
$I_{it}r^e_{nSP,t}$	-0.150**	0.067	-0.194***	0.036	0.057	-0.029	-0.180**	0.011	-0.160**
	(0.063)	(0.041)	(0.061)	(0.066)	(0.039)	(0.064)	(0.083)	(0.022)	(0.067)
Observations	11482	11482	11482	11463	11463	11463	11463	11463	11463
R-squared	0.339	0.663	0.042	0.324	0.670	0.028	0.014	0.004	0.014
1991-2005	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$	$r^e_{i,t}$	$-N_{iDR,t}$	$N_{iCF,t}$
$r^e_{SP,t}$	0.584***	0.534***	0.016	0.632***	0.568***	0.028	-0.049	-0.035	-0.011
	(0.121)	(0.084)	(0.152)	(0.101)	(0.088)	(0.137)	(0.074)	(0.022)	(0.073)
$I_{it}r^e_{SP,t}$	0.533***	0.090*	0.393***	0.448***	0.084	0.314***	0.084	0.007	0.075
	(0.111)	(0.049)	(0.097)	(0.113)	(0.060)	(0.110)	(0.129)	(0.036)	(0.118)
$r^e_{nSP,t}$	0.477***	0.319***	0.062	0.343***	0.313***	-0.040	0.135**	0.007	0.103*
	(0.091)	(0.070)	(0.108)	(0.084)	(0.074)	(0.107)	(0.060)	(0.017)	(0.056)
$I_{it}r^e_{nSP,t}$	-0.325***	0.002	-0.260***	-0.234**	0.001	-0.182**	-0.092	0.001	-0.078
	(0.087)	(0.036)	(0.075)	(0.090)	(0.042)	(0.089)	(0.089)	(0.027)	(0.082)
Observations	9636	9636	9636	9603	9603	9603	9603	9603	9603
R-squared	0.188	0.556	0.018	0.175	0.559	0.012	0.012	0.005	0.010

Table 8: **Robustness to growth-value subsamples**

This table follows the same sample and procedure as table 6. Panel A shows the results using only the bottom quintile of event stocks sorted on book-to-market ratios. Panel B shows the results using only the top quintile of event stocks sorted on book-to-market ratios. The equations estimated are similar to those in table 6. I include firm dummies, and the standard errors are clustered by time to account for cross-sectional autocorrelation.

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

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

Table 9: **Second Order Aggregate VAR**

This table shows the OLS parameter estimates for a second-order monthly aggregate VAR model including a constant, the log excess market return (r^e_M), the term yield spread (TY), the log price-earnings ratio (PE), and the small-stock value spread (VS). Each set of two rows corresponds to a different dependent variable. The first nine columns report coefficients on the nine explanatory variables and the tenth column reports the corresponding adjusted R^2. Standard errors are in parentheses. The sample period for the dependent variables is January 1929 - May 2009, providing 965 monthly data points.

Firm-level VAR for excess returns

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

Table 10: **Second Order Firm-level VAR**

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

Second order firm-level VAR for market-adjusted returns

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

Table 11: **Changes in beta and matching firms for a second-order VAR**

This table shows the changes in the slope of regressions of returns (and its components extracted from a second-order VAR) of stocks added to the S&P 500, matched stocks, and their difference, on returns of the S&P 500 index and the non-S&P 500 rest of the market. Firms are matched to event stocks based on industry and size, as described in the text. The sample and definition of variables is described in Table 3. The equations estimated are the following:

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

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

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

Table 12: **Alternative VAR: different specifications**

This table shows the pooled-WLS parameter estimates for the first equation of a first-order monthly firm-level VAR model. The state variables include a constant, a set of firm-level variables, and two sets of aggregate variables. The firm level variables are: the excess log stock return ($r^e_{i,t}$), stock momentum ($MOM^e_{i,t}$), the log book-to-market ratio ($BM_{i,t}$), and the log profitability in excess of the risk free rate. The first set of aggregate variables is formed by the cross-sectional median of each of the firm-level variables. The second set of aggregate variables consists of the log excess market return (r^e_M), the term yield spread (TY), the log price-earnings ratio (PE), and the small-stock value spread (VS). Standard errors are in parentheses. The weights used in the WLS estimation are proportional to the inverse of the number of stocks in the corresponding cross section. Standard errors (in parentheses) take into account clustering in each cross section. The sample period for the dependent variables is January 1954 - December 2008, providing 660 monthly cross-sections and 1,658,049 firm-months.

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

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

Table 13: **Alternative VAR: predicting firm-level excess returns**

This table shows the pooled-WLS parameter estimates for all the equations in the first-order monthly firm-level VAR model corresponding to column (4) in the previous Table. The model state vector includes a set of firm-level variables: the excess log stock return ($r^e_{i,t}$), stock momentum ($MOM^e_{i,t}$), the log book-to-market ratio ($BM_{i,t}$), and the log profitability in excess of the risk free rate; and a set of aggregate variables: log excess market return (r^e_M), and the term yield spread (TY). Rows corresponds to dependent variables and columns to independent (lagged dependent) variables. The first seven columns report coefficients on the seven explanatory variables and the eighth column reports the corresponding adjusted R^2. The weights used in the WLS estimation are proportional to the inverse of the number of stocks in the corresponding cross section. Standard errors (in parentheses) take into account clustering in each cross section. The sample period for the dependent variables is January 1954 - December 2008, providing 660 monthly cross-sections and 1,658,049 firm-months.

Firm-level VAR for excess returns

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

Table 14: **Changes in beta and matching firms for an alternative VAR**

This table shows the changes in the slope of regressions of returns (and its components) of stocks added to the S&P 500, matched stocks, and their difference, on returns of the S&P 500 index and the non-S&P 500 rest of the market. Firms are matched to event stocks based on industry and size, as described in the text. The sample and definition of variables is described in Table 3. The equations estimated are the following:

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

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

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

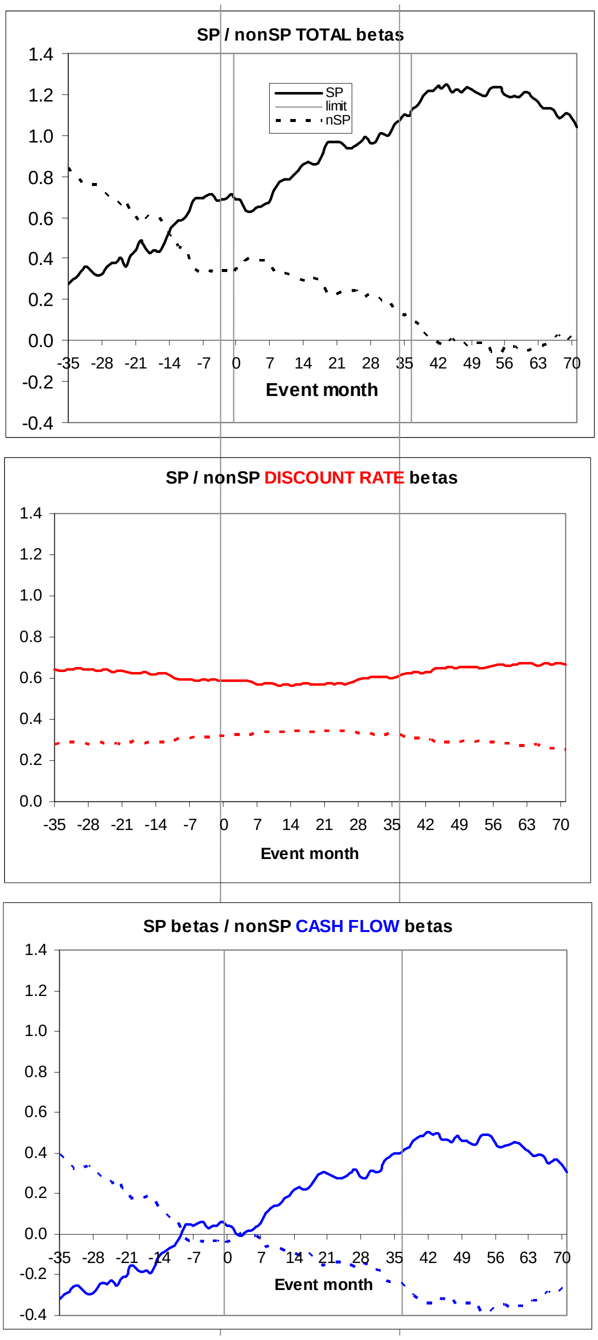

Figure 1: This figure plots the evolutions of betas around S&P 500 inclusions. In the left Panel, I plot the evolution of the overall beta and in the right Panel the two different components of beta are displayed.

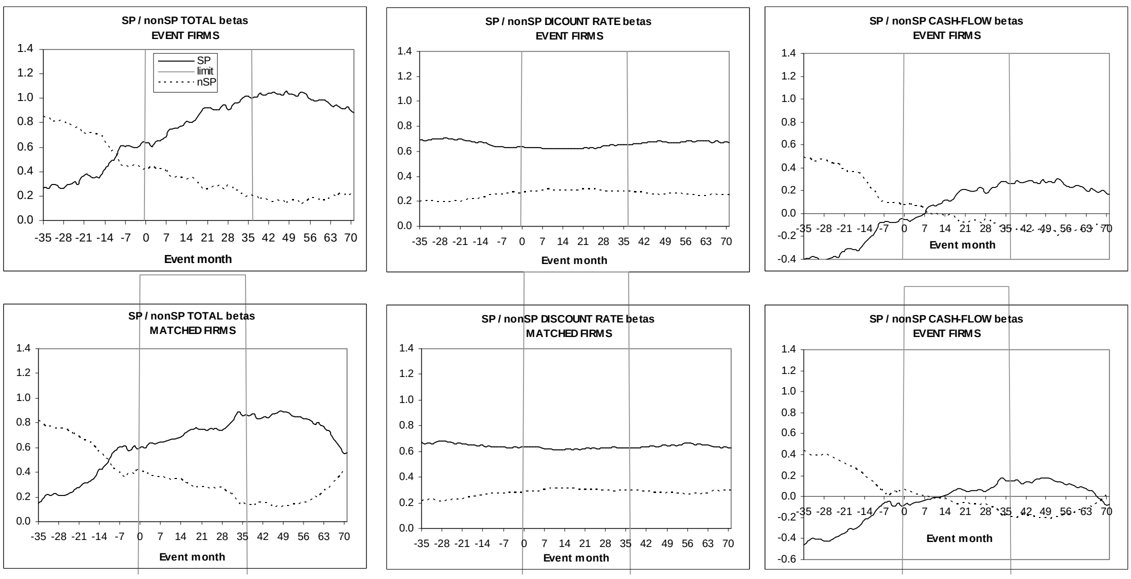

Figure 2: This figure shows rolling betas (total, discount rate, and cash-flow betas), for event stocks (top panel), and matched stocks (bottom panel).