The Growth of Index Investing & Financial Markets

1 Introduction

Index investing has seen dramatic growth over the past 40 years. Assets under Management (AUM) in index funds have increased from $511 million in 1985, to $55 billion in 1995, to over $4.0 trillion today. After including index-linked Exchange Traded Funds (ETFs), the total value of indexed assets now surpasses $5.5 trillion. Rather than due to a market-wide growth in assets, where a "rising tide lifts all boats", growth in passive funds has come largely at the expense of traditional active managers. Actively managed mutual funds have experienced outflows of over $1 trillion over the past ten years, and hedge fund closures have outnumbered openings for the first time since the financial crisis. Even well-established hedge funds with "superstar" managers have turned into family offices (e.g. BlueCrest, Seneca Capital, Soros Fund Management) or have shut down in recent years (e.g. Eton Park, Perry Capital, and Chesapeake Partners). With an ever-increasing amount of capital shifting into index funds and ETFs from active managers, it is important to consider the impact of these strategies on the financial markets and the real economy. To this point, the economic impact of index investing leaves a number of questions. In this paper, I evaluate whether this substantive shift from active to passive investing has altered key characteristics of equity securities. More specifically, I examine the relationship between capital flows into index-tracking vehicles and effects on price levels, comovement, and liquidity characteristics of index constituents relative to out-of-index peers. From a broader perspective, I attempt to answer the following question: is the growth of index investing something to celebrate unambiguously or are there potentially negative side effects?

Stock indexes provide a convenient way to distill the performance of a basket of securities into one easily reportable number. Statements like "The S&P 500 is up .54% today on a positive jobs report" allow market participants to absorb important information about the general direction of sentiment, performance, and the

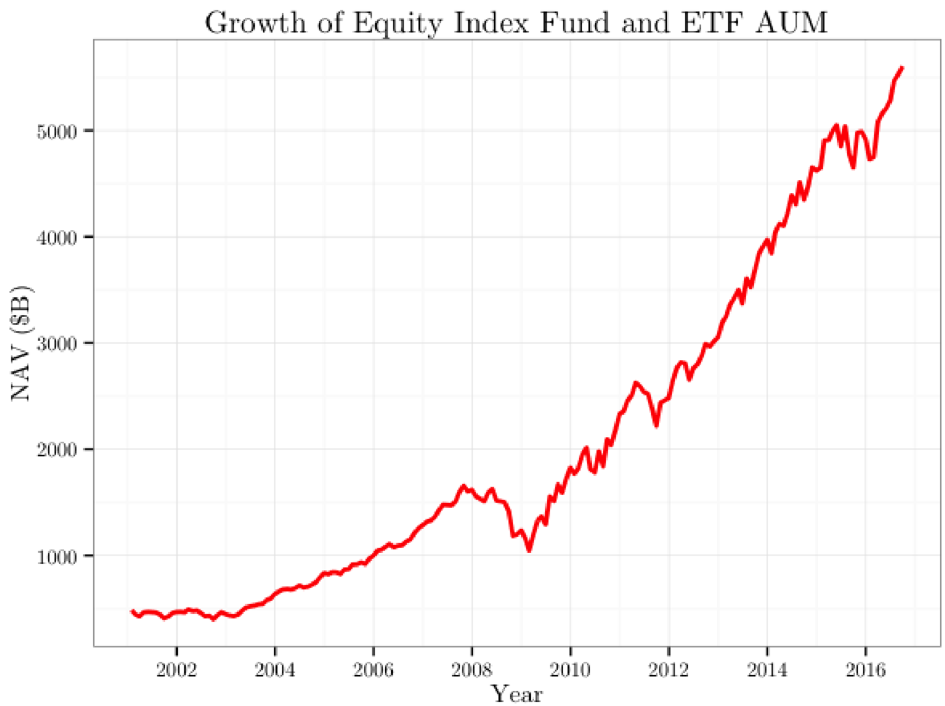

Figure 1: Index Funds and ETF AUM has grown nearly 1200% over the last 15 years. Data on passive fund AUM comes from CRSP Mutual Fund database. The sample includes all funds flagged as index funds or ETFs.

impact of economic news. Stock indexes have expanded over the years to summarize performance of companies in individual sectors, nations, or with similar security characteristics (e.g. market capitalization, dividend yield, or historical volatility). Broad indexes such as the S&P 500, Dow Jones Industrial Average, and the Nasdaq Composite Index—which include stocks covering a large percentage of total market capitalization—are frequently used to assess investors' views on economic health, growth prospects, and overall market valuation. To construct an index, an agency such as Russell Investments or Standard & Poor's aggregates a group of stocks and computes a weighted average of their prices. Indexes can either be capitalization-weighted, where each stock's contribution to the index is weighted by its market capitalization (such as the S&P 500 or Russell 3000), or price-weighted, where each stock's contribution to the index is weighted by its stock price (such as the Dow Jones

Industrial Average or Nikkei 225). Index values are updated in real time based on the prices of its components. Component weightings are updated on either a quarterly, semiannual, or annual basis. Originally summarizing the performance of the largest stocks (Charles Dow's Dow Jones Industrial Average originally included just the 12 largest US corporations), indexes now exist for niches as narrow as Zimbabwean mining corporations, Nepali hydro power companies, and Sri Lankan footwear and textile producers. From the first index reported by Charles Dow in 1896 to the 53 U.S. stock indexes reported every day in the Wall Street Journal, indexing has revolutionized the means by and detail to which investors track financial markets.

As the most visible indicators of sentiment and performance in the financial markets, naturally investors have begun to benchmark themselves to and invest in securities indirectly or directly related to stock indexes. To more precisely define index investing, I offer the following definition: investing in a predefined and publicly known list of stocks using a publicly disclosed and replicable strategy[1]. Under this definition, both index funds[2] and ETFs, whose publicly traded 'shares' can be exchanged at any time for the underlying security basket, are considered forms of index investing.

Index investing notably differs from the traditional active management model. Unlike index investors, who hold all securities in their tracked index, active managers seek to take advantage of mispriced securities, buying securities that are undervalued and short-selling securities that are overvalued. Active managers use expertise in assessing company management, industry health, cash flows, and growth potential to identify mispriced stocks. Due to constrained resources and the high cost of acquiring enough information to form a developed opinion of a company (in terms of time, compensation for analysts, cost of data acquisition), active managers typically hold only a small fraction of companies in their benchmark index. Investors in actively

1. This definition mirrors Wurgler (2011) with one notable difference, the addition of public disclosure of strategy.

2. Index funds are defined as open or closed-end mutual funds that hold index constituents in the same weighting as the underlying index.

managed funds believe that the expertise of active managers in identifying mispriced securities will result in returns exceeding those of the underlying index.

One major argument for index investing over active management is that prices are efficient, meaning that it makes no sense to pay active managers high fees and commissions. In its semi-strong form, the most commonly accepted by financial economists, the Efficient Markets Hypothesis states that security prices incorporate all public information and that neither fundamental analysis nor technical analysis can result in consistent excess returns. Proponents of this view point to the considerable body of evidence showing that mutual funds underperform their benchmark indexes once fees are taken into account and that there is relatively low persistence in performance for active funds[3]. Investing in a passive vehicle implicitly accepts that "beating" the market represents luck, not skill, and that receiving market return and paying (usually) significantly lower fund fees represents a superior form of investment management. The view that active management does not generate enough value to justify fees extends beyond academia. Even legendary active money manager Warren Buffett has repeatedly claimed that the entire active management industry exists to siphon fees without delivering real outperformance[4].

Given the expansive literature on the underperformance of active management and increased transparency regarding fund performance, investing in a portfolio of index funds has quickly become the default financial advice given to retail investors. Many of the new dollars invested in the global equity and fixed income markets have flocked towards index funds and ETFs, totalling over $3 trillion since 2001. Over the past ten years, much of this growth has come at the expense of active managers, who have experienced outflows of over $1 trillion over the same period, as seen in Figure 2 below. The vast majority of outflows from active funds have come in the last seven

3. See Carhart (1997), Malkiel (1995), Gruber (1996), Wermers (1999), Petajisto (2013) and others.

4. See 2016 Berkshire Hathaway Letter to Shareholders.

years since the financial crisis.

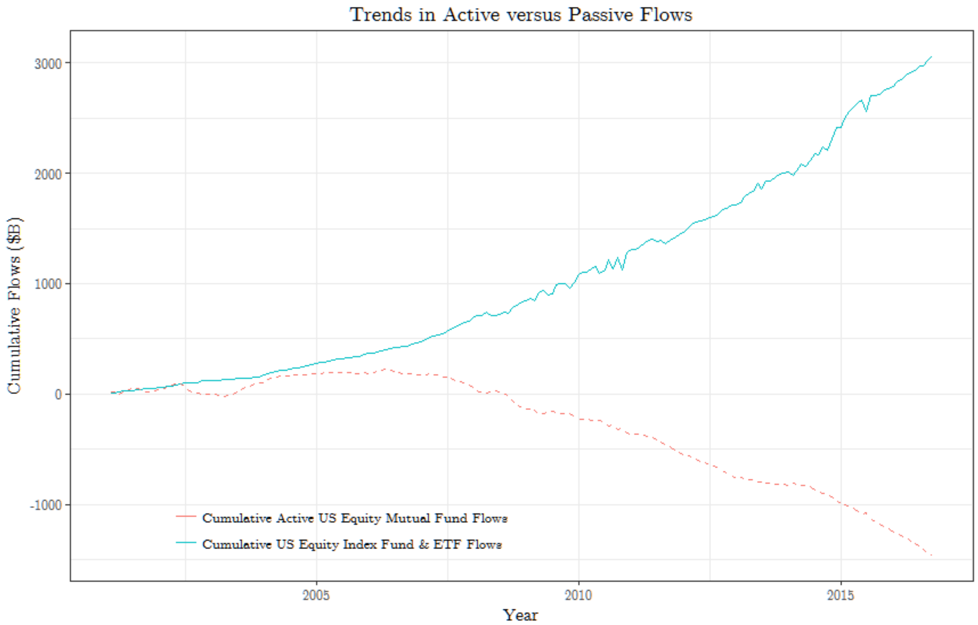

Figure 2: Growth of Passive Investing at the Expense of Active Managers.
Data on passive fund AUM comes from CRSP Mutual Fund database. The sample includes all funds flagged as index funds or ETFs. Data on active flows comes from the Investment Company Institute, and includes all U.S. focused mutual funds. I remove the impact of passive funds by subtracting all flows into index funds. I merge the two data sets by month and calculate cumulative flows.

In addition to outflows from actively managed mutual funds, mutual fund management fees have largely compressed over the past fifteen years. The average annual expense ratio (measured as a percentage of assets per year paid to the fund manager as a management fee) of U.S. based equity mutual funds has slid from 99 basis points in 2000 to just 68 basis points in 2015[5], partially attributable to competition from index funds offering fees as low as 5 basis points (in the case of the Vanguard S&P 500 ETF). The top seven ETFs and mutual funds by AUM are currently all index funds, and each boasts annual fees below 0.09%. Of the seven largest actively managed ETFs and mutual funds, none offers an annual fee less than *six times* that amount. In 2004, only one of the top seven funds was an index fund,

5. See Table 3 for a detailed overview on the evolution of equity fund fees.

and fund fees for the top active funds often exceeded 1%.

Another impact of the growing popularity and visibility of indexes and the proliferation of index funds on active managers is a phenomenon known as "shadow indexing", where active managers who benchmark performance to an index hold securities that mimic index components. To illustrate an example of shadow indexing, consider the following case: a long/short equity hedge fund manager is benchmarked against the S&P 500, meaning that any performance-driven fees she receives come from outperforming the S&P 500. In Q1, she delivers 10% returns in excess of the S&P 500 (α of 10%). Happy with her 10% outperformance, she may exit her positions and merely replicate the index for the rest of the year in order to "lock in" her fees for the year. While technically the fund is still an actively managed fund, for these next three quarters the fund serves essentially as an index fund. Though this is perhaps an extreme case, consciously or unconsciously, active managers tend to mirror the performance of their benchmark index. Miller (2006) finds that the average mutual fund has over 90% of its return variance explained by its benchmark index. Including shadow indexing, it is tough to even quantify the size of the impact that index-linked investing has had on the broader equity markets. Index-linked investing has redefined the investment landscape, changing the way that active managers deploy capital and driving investors towards lower-fee, passive options.

In this paper, I analyze the impact of the growth of index investing on security characteristics in the United States stock market. I consider three major areas of impact. First, I evaluate the relationship between the growth of index-linked assets under management and price levels. Many practitioners, perhaps most notably Bill Ackman, founder and CEO of Pershing Square Capital Management, a $12.4 billion long-short equity hedge fund, have discussed potential price distorting effects of growing indexed assets. In Ackman's 2015 Annual Letter to Shareholders, he specifically targets index funds in what he describes as the "index fund bubble."

Centrally, Ackman states that index funds are inherently momentum investors who are forced to hold more shares of companies that have performed well, unlike value-oriented active managers. Coupled with the assertion that valuations in the short term are driven by capital flows (overwhelmingly positive in the recent years for index funds), Ackman asserts that index fund demand has created artificially high prices for index components, most notably in the S&P 500. In the long run, where valuations converge to intrinsic value per textbook asset pricing theory, the rate of return for index components will decrease, resulting in index fund under-performance and index fund outflows (i.e. the bubble will burst). I test whether the growth of indexation has affected the prices or relative valuation of index components within popular indexes such as the S&P 500.

Second, I ask whether increased dollars in passive vehicles affect patterns of volatility and comovement in stocks entering or exiting a heavily-traded index. A large body of research discusses the change in security comovement patterns upon inclusion to and deletion from heavily traded indexes[6]. I extend this research to analyze whether the size of the inclusion/deletion change depends on the amount of assets benchmarked to the index, the size of inflows into funds tracking the index, or the volatility of index-tracking AUM.

Third, I examine the impact of the growth of index investing on the liquidity of stocks within the index. When stocks are added to a heavily-traded index such as the S&P 500, one would expect that they would become much more liquid, given the increased exposure to index-constrained mutual funds and index tracking products. In this paper, I evaluate whether this phenomenon exists and, if so, whether the size of changes in security liquidity upon index inclusion and deletion depends on the amount of assets in passive funds or flows into passive funds over the event period.

Overall, I find that while index investing has some effects on each of the

6. See Barberis, Shleifer, and Wurgler (2005), Kasch and Sarkar (2011), Coakley and Kougoulis (2004), Greenwood and Sosner (2007), among others.

security characteristics (price level, comovement, liquidity) I test, the magnitude of these effects—up to this point—have remained constant over the sample period despite huge growth in passive AUM. Specifically, I observe limited evidence of an index fund bubble in security pricing and that both comovement and liquidity effects upon index inclusion/deletion are time-invariant. Perhaps the reason for the lack of effects lies in the still limited share of passive fund ownership in large indexes such as the S&P 500. Alternately, the effects of index investing on these characteristics may appear more strongly in smaller indexes where index funds and ETFs represent a larger percentage of stock ownership and trading volume. In all, my findings indicate that the growth of index investing has not resulted in significant changes that should overly concern economists or practitioners.

I structure my paper as follows: Section 2 describes the data. Section 3 discusses the impact of the shift towards index investing on price levels and detachment. Section 4 examines the volatility and comovement effects of index inclusion and deletion and the relationship of these effects to increased passive dollars. Section 5 summarizes the effects of index inclusion and deletion and the size of passive fund AUM on stock liquidity and price informativeness. Section 6 offers some concluding remarks.

2 Data Sources

To evaluate the relationship between flows and various security characteristics, I first need to construct a data set containing information on flows into and out of index funds. To build this data set, I require information on fund net asset values and returns, given that to my knowledge no existing comprehensive data set of index fund and ETF flows exists. With this data, a set of flows can be created fairly simply. I use monthly data on net asset values and returns for index funds and ETFs from the

Center for Research in Security Prices (CRSP). CRSP at the University of Chicago is a leading research data provider for security pricing, mutual funds, and stock index components and pricing. CRSP mutual fund data is available on a monthly basis from December 1961 to present, and includes data from all publicly traded mutual funds.

To measure assets tied to indexes, I focus on two primary financial instruments: ETFs and index funds, both within the United States. Particularly, I focus on funds tracking the S&P 500 Index. These funds are typically the largest funds by AUM and are particularly important to retail investors. Per the Investment Company Institute, approximately 31% of index fund assets follow the S&P 500 Index as of 2016, nearly as much as every other US domestic equity index combined. To construct this sample, I pull all funds from the CRSP Mutual Fund Names File flagged as ETFs or index funds, filter for all mentions of "S&P 500" or "500" within the *fund_name* field, and exclude all sector-specific, short-biased, or leveraged funds on an ad hoc basis[7]. This method results in a total sample of 274 funds. These include three of the largest ten mutual funds and ETFs by AUM, including the SPDR S&P 500 ETF (Ticker: SPY), the Vanguard 500 Index Admiral Class Shares (Ticker: VFIAX), and the iShares Core S&P 500 Index ETF (Ticker: IVV). Collectively, as of September 2016, these 274 funds manage in excess of USD $840 billion of investors' assets.

CRSP provides data on both a monthly and daily time scale for both returns and fund Net Asset Value (NAV). Given the $t+1$ reporting bias as discussed by Staer (2016) and Quinn et al. (2006), in which shares outstanding are reported with a one day lag by mutual funds and ETFs, along with the relative frequency of ETF zero-flow days in which there are no redemption or creation activities, monthly data provides a more accurate picture of the relationship between flows and returns. Zero-flow days

7. The removal process excludes funds such as the Rydex Inverse S&P 500 fund, Direxion Daily S&P 500 Bull 2X Shares, and the SPDR S&P Biotech ETF. I exclude these funds because they either do not affect the entire index or use derivative securities and shorting as opposed to positions in the underlying stock.

are frequent, especially in smaller funds, due to limited daily creation-redemption activity. For this reason, I choose to use monthly data on flows and returns.

To isolate the impact of flows into passive funds, I also require data on flows into actively managed funds. To control for fund flows to actively managed funds and passive funds which do not track the S&P 500, I use aggregate mutual fund flows data provided by the Investment Company Institute (ICI). Every month, ICI provides aggregate flows into mutual funds in each asset class. The Institute is a global trade association for mutual funds, ETFs, and unit investment trusts representing over \$18.2 trillion of AUM, consisting of nearly every U.S. mutual fund provider. The data is separated into 13 different classes based on fund strategy and objective, into categories such as sector equity, aggressive growth, and capital appreciation. This data is complete from January 1984 to present, and limited data for some categories exists for the period from 1976-1984.

Based on the length of fund history provided by CRSP and the ICI, my sample begins in August 1976 and extends until September 2016. However, since CRSP only provides mutual fund data on a quarterly basis until January 1989, all regression analysis will be limited to data from January 1989 to September 2016. Given the rapid growth of index funds and ETFs over the past decade, I subdivide my sample into the following sub-periods to analyze the flow-return relationship over different points in the cycle: 1989-2000, 2001-2010, and 2011-2016. To analyze whether this flow-return relationship varies during negative-return shocks, I additionally add a subsample from December 2007 to June 2009, the period of the Global Financial Crisis as defined by the NBER. Figures 3 and 4, below, illustrate the growth of assets managed by the the 274 S&P 500 tracking funds in my sample and the growth in the percentage of total index capitalization owned by these funds. It is important to note that the true passive ownership of the S&P 500 likely exceeds the 4.5% peak documented in Figure 4. Given the construction of my sample, I isolate index fund and ETF flows directed solely towards the S&P 500. Many additional index funds and ETFs track broader indexes that include most or all of the securities within the S&P 500, such as the Russell 3000, the Wilshire 5000, and the Dow Jones Total Stock Market Index. Most estimates of true passive ownership within the S&P 500 fall within the 10-15% range, with the Wall Street Journal recently estimating total ownership at 11.6%[8]. Table 1, Panel A provides summary statistics on the relative magnitude of S&P 500 index fund and ETF flows.

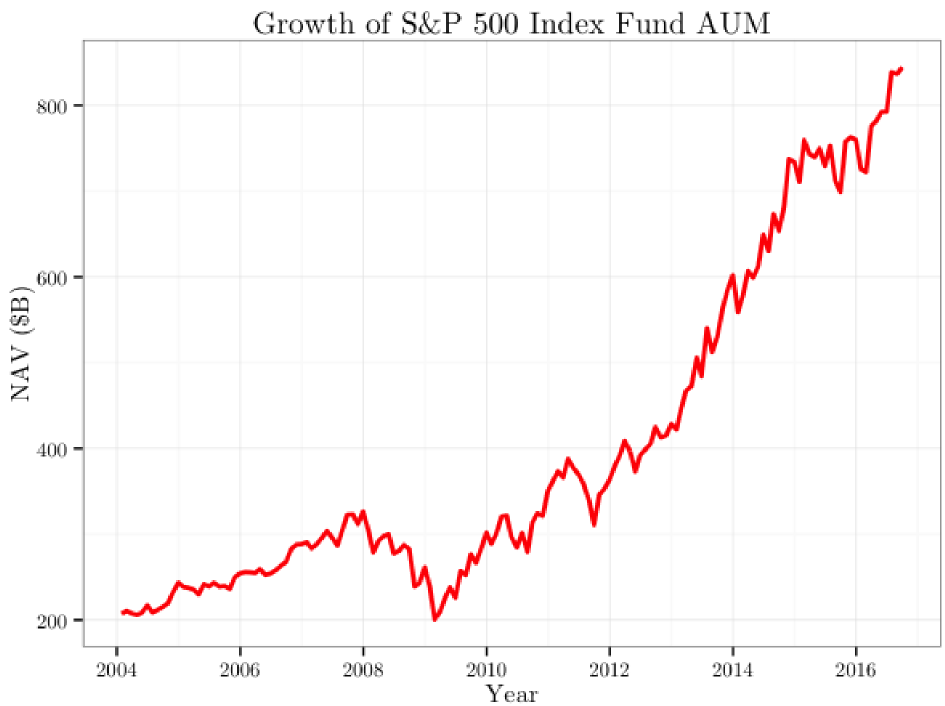

Figure 3: Growth of S&P 500 Tracking Funds from 1976-2015.
Data on passive fund AUM comes from CRSP Mutual Fund database. The sample includes all funds flagged as index funds or ETFs mentioning "500" or "S&P 500" in their investment objectives or fund name excluding leveraged or inverse ETFs.

8. McGinty, Tom, Sarah Krouse, and Elliot Bentley. "Index Funds Are Taking Over the S&P 500." The Wall Street Journal, October 2016.

Figure 4: Growth of Passive S&P 500 Ownership Share from 1976-2015.
Data on passive fund AUM comes from CRSP Mutual Fund database. The sample includes all funds flagged as index funds or ETFs mentioning "500" or "S&P 500" in their investment objectives or fund name excluding levered or inverse ETFs. S&P 500 total market capitalization comes from CRSP Monthly S&P Universe file.

2.1 Flow Variable Construction

Following the literature, including Staer (2016), Kasch and Sarkar (2011), Barberis, Shleifer and Wurgler (2005), I gather data on index returns from the CRSP S&P 500 Index File and data on fund net asset value and returns from the CRSP Mutual Funds database. From this data, beginning in January 1976 and ending in June 2016, I construct a monthly fund flow metric using the following methodology:

$$f_t = \frac{NAV_{f,t}}{1 + r_{t-1}} - NAV_{f,t-1} \tag{1}$$

I define r_{t-1} as the return over the period from $t-1$ to t and $NAV_{f,t}$ as the Net Asset Value of the fund f at time t. To normalize the flows and get a metric for the relative size of each flow, at each time period t, I divide my fund flow metric by the total market capitalization of the underlying index at time t (referred to as $IndexCap_t$). My final metric for monthly flows is the following:

$$f_t = \frac{\frac{NAV_{f,t}}{(1+r_{t-1}}) - NAV_{f,t-1}}{IndexCap_t} \tag{2}$$

To come up with an aggregate measure of flows, I sum each individual flow from each fund with available data during the given time period, as follows (in regressions, I refer to this metric as Scaled Flows):

$$F_t = \sum_{f \in Funds} f_t = \sum_{f \in Funds} \frac{\frac{NAV_{f,t}}{(1+r_{t-1}}) - NAV_{f,t-1}}{IndexCap_t} \tag{3}$$

Following Goetzmann and Massa (2003), who additionally adjust flows to account for idiosyncratic fund growth through adjusting by the 6-month trailing moving average flow rate, I construct a secondary variable, which I refer to as Adjusted Flows in all regressions. This method seeks to isolate the unexpected component of fund flows. Mathematically, this metric can be represented as the following:

$$AF_t = F_t - \sum_{i=t-6}^{t} \frac{F_i}{6} \tag{4}$$

As a robustness check, I create a variant of this metric adjusting solely for previous period flow ($F_t - F_{t-1}$) and a metric using the past three period flows in an AR(3) framework following Warther (1995) and rerun all analyses. I find no variation in using these metrics versus the adjusted flows metric above.

To isolate the impact of index investing, I create controls for active fund flows to include in all regression analysis. To control for active mutual fund flows, I create a Scaled Active Flows metric using the identical methodology that I use for index fund flows. To capture domestic equity fund flows, I aggregate flows for each time period for funds classified as aggressive growth, growth and income, growth, sector equity, income equity, and regional equity strategy funds by the ICI. These constitute seven of the ten ICI categories for equity mutual funds. I exclude funds categorized as international equity, global equity, or emerging markets funds due to non-US exposure. To isolate flows not coming from S&P 500 index funds, I then subtract aggregate S&P 500 index fund flows from this aggregate measure, since index fund flows are included in the ICI dataset. Since ETFs are excluded from the ICI dataset, I only subtract flows from index funds. I scale these flows by total S&P 500 index capitalization in order to preserve similarity between the two flow metrics. In order to preserve symmetry, I create an Adjusted Active Flows metric by adjusting for the 6-month trailing moving average flow rate to isolate the unexpected component of active mutual fund flows.

2.2 Summary Statistics

Basic descriptions of the data can be found in Panel A of Table 1, which summarizes net flows, returns, and NAV of funds in my entire sample. Over the entire sample, I have 333 monthly data points. Average monthly returns for all S&P 500 tracking funds are between .45 and .48 basis points, corresponding to slightly over 5% annualized returns, slightly higher than the S&P 500 returned over the period. This can be attributed to the reinvestment of dividends, index tracking error due to fund expenses such as trading costs, and other activities such as securities lending. The most negative monthly return (of approximately -16.9%) comes in October 2008, while the most positive return (of approximately 11.2%) comes in October 2011. Notably, over the whole sample, inflows to S&P 500 funds average just over $1 billion per month. When scaled to correspond to a percentage of the total S&P 500 index capitalization each day, flows to each fund are quite small. For my aggregate S&P 500 fund flows measure, the greatest positive flow shocks reached approximately 0.22% of total

S&P500 index capitalization, with a maximum draw-down of about 0.21%. Average flows to S&P500 funds were trivial as a percentage of total index capitalization, but have grown in magnitude over the sample.

Interestingly, the correlations between S&P 500 fund flows across different funds are fairly small and insignificant over all time horizons (see Table 2 for a summary of the three largest funds). If flows were driven by returns chasing behavior, one would expect to see high correlations between inflows and outflows of each fund as each tracks the same underlying index. In fact, in the financial crisis, when one would expect return chasers to pull their money out of funds, nearly all funds recorded positive average inflows and flows still had relatively insignificant correlations with each other.

3 Level of Prices

In this section, I analyze the impact of increasing index fund and ETF assets under management on the level of prices of equity securities within and outside of the index. More specifically, I examine the relationship between S&P 500 index-tracking vehicle flows and underlying index return over the short-term and long-term and the effect of these flows on the relative performance and valuation of securities within the S&P 500 versus close peers. An immediate concern regarding the growth of index investing relates to the effect of flows on returns. With billions of dollars of net flows every month—sometimes even tens of billions—it seems altogether reasonable for flows to asymmetrically affect the prices of index components. Most of the literature around indexation has discussed the presence, persistence, and causes of these effects. Much of this work focuses on event studies on the returns of stocks added to or removed from major indexes. Within the S&P 500, there is considerable evidence of index inclusion effects, in which stocks added to the S&P 500 index see immediate and persistent price increases. The reverse holds true for companies removed from the S&P 500, which see immediate price declines.

To contextualize any S&P 500 index inclusion and deletion effects, one must consider Standard & Poor's' (the firm which manages the index) methodology governing index changes. Compared to other major indexes such as the Russell 2000 or Dow Jones Industrial Average, the S&P 500 has comparatively subjective inclusion and exclusion requirements. This allows for a more precise test of inclusion and deletion effects because it is harder for arbitrageurs to front-run the index change and buy stocks that appear to be prime candidates for inclusion and short those that are likely to be removed. To be a candidate for inclusion in the S&P 500, a stock has to meet various criteria that are updated on an annual basis, including a minimum market capitalization ($5.9 billion as of most recent guidelines), a minimum liquidity threshold based on the percentage of public float[9] and trading volume, positive GAAP LTM earnings-per-share, and U.S. domicile. For companies within the index, removals occur when corporate actions including financial restructurings, mergers and acquisitions (M&A) or bankruptcy result in the company violating inclusion criteria or if the company itself substantially violates inclusion criteria. However these changes are not mechanical, and S&P reserves discretion upon deciding additions and deletions. In fact, "S&P Dow Jones Indices believes turnover in index membership should be avoided when possible"[10]. Index reconstitution occurs on an ad-hoc basis at irregular intervals. Changes as a result of M&A typically occur on the effective date of transactions for companies within the index and compose the bulk of index changes, while market capitalization-driven updates occur more infrequently[11]. Mechanically, within the S&P 500, it becomes very difficult for potential arbitrageurs to

9. Public float indicates the percentage of shares that are available to trade on the public markets (excludes shares held by company officers, insiders, et cetera).

10. From February 2017 *S&P Dow Jones Indices: Index Methodology.*

11. The most recent change to the S&P 500 due to market capitalization was on December 2, 2016, while three M&A related changes have happened in the first week of March

predict the exact timing of index inclusions or deletions and front run them if these index effects do exist.

Within the S&P 500, Shleifer (1986) documents significant positive post-inclusion return effects and attributes them to a rightward shift in the demand curve for the security. These effects have been replicated by many extensions, including Wurgler and Zhuravskaya (2002), Lynch and Mendenhall (1997), Jain (1987), Hegde and McDermott (2003) and others. There are four common explanations of these phenomena: (i) a rightward shift in demand (Shleifer 1986); (ii) an increase in expected cash flows as inclusion in the index is considered positive sentiment about the company (Jain 1987); (iii) a rise in investor awareness of the security (Chen et al. 2004); (iv) a decreased discount rate corresponding to increased liquidity (Hegde and McDermott 2003). Apart from event studies of inclusion events, some research extends similar supply and demand explanations of security valuation to the broader market. Warther (1995) documents a relationship between mutual fund flows and stock prices. Unexpected flows (defined as the residual of an AR(3) model based on past flows) into mutual funds result in a sharp increase in the prices of assets the fund invests in. This unexpected demand component shifts the demand curve right and thus increases prices. Rather than due to return-chasing behavior, he argues that since his regressions do not show a relationship between lagged returns and unexpected fund flows, this effect demonstrates some causality (although limited by only monthly flows and returns data). In this paper, I use a modified version of Warther's unexpected flow metric as one explanatory variable. Morck and Yang (2011) document a significant valuation premium of 40% for S&P 500 components versus matched non-index companies with similar industry and size classifications and attribute this largely to supply and demand effects.

The impact of supply and demand specifically coming from index funds and ETFs on the price or required rate of return of index components outside of inclusion and deletion effects has not been substantially covered by economists. An exception is work by Goetzmann and Massa (2003), who have observed a strong negative relationship between index fund outflows and S&P 500 returns, along with a positive relationship between index fund inflows and S&P 500 futures returns. Additionally, using daily data, they see no immediate correction to these demand curve shifts, implying that these return effects are permanent. In this section, I will connect index fund and ETF supply and demand with aggregate index price levels by determining the effect of index fund flows on contemporary and future returns and evaluating the effect of index flows on the relative valuations of index components versus out of index components. I build upon previous literature concerning index supply/demand effects and detachment effects including Morck and Yang (2011), Goetzmann and Massa (2003), and Warther (1995).

3.1 Passive Flows and Contemporaneous Index Returns

To analyze the contemporaneous relationship between flows and returns, I run OLS regressions between passive S&P 500 flows and the contemporaneous return on underlying index. These regressions are of the form:

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 F_t \tag{5}$$

I run these regressions for the cumulative flows across all 274 funds. Additionally, I run regressions following Goetzmann and Massa (2003) and adjust for the 180-day moving average of fund flows, to eliminate fund-specific idiosyncrasies and isolate an unexpected component of flows. I report the results of these regressions in Table 4. Mathematically, these regressions are of the form:

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 AF_t = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6}) \tag{6}$$

In Table 4b, additional regressions including controls for aggregate mutual fund flows are shown. These regressions take the form:

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 F_t + \hat{\beta}_2 ActiveFlow_t \tag{7}$$

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6}) + \hat{\beta}_2 (ActiveFlow_t - \sum_{j=t-6}^{t} \frac{ActiveFlow_j}{6}) \tag{8}$$

To further isolate a causative link between flows and underlying returns, I additionally analyze the correlation between lagged returns and flows, to evaluate whether flows chase returns. I evaluate the results of these regressions for three time horizons, from $t-3$ to $t-1$. Mathematically, these regressions can be expressed as:

$$\hat{F}_t = \hat{\beta}_0 + \hat{\beta}_1 r_{t-1} + ... + \hat{\beta}_{i-1} r_{t-(i-1)} + \hat{\beta}_i r_{t-i} \quad i \in (1,2,3) \tag{9}$$

$$\hat{AF}_t = \hat{\beta}_0 + \hat{\beta}_1 r_{t-1} + ... + \hat{\beta}_{i-1} r_{t-(i-1)} + \hat{\beta}_i r_{t-i} \quad i \in (1,2,3) \tag{10}$$

The results of these regressions are shown in Table 4c. These regressions test to determine whether flows increase following periods of high return and display return-chasing behavior.

Regression (1) in Table 4b indicates that the scaled level of index fund flows does not have a statistically significant relationship with underlying index return. However, once I add the control variable denoting scaled active fund flows during the same period in Regression (3), both types of flows demonstrate a strong positive relationship with underlying return. The Adjusted R^2 of this regression remains fairly low at .123, although this is to be expected as index returns should largely be driven by fundamental factors outside of flows.

I additionally isolate the unexpected portion of flows in Regression (2) in

Table 4b. The results indicate a strong positive relationship between unexpected index fund flows and underlying index performance even without adjusting for active fund flows. While the Adjusted R^2 of this regression is very low, at just 0.018, once I add a control for unexpected scaled active fund flows in Regression (4), the Adjusted R^2 climbs up to a significant .257. In this regression, both unexpected index fund flows and unexpected active flows are very significant, with $t-$values of 7.1 and 10.0 respectively. The sign of the relationship between fund flows and returns is positive in all my regressions, indicating that in periods of high inflows, the underlying index tends to outperform. These results align with both those of Goetzmann and Massa (2003) and Warther (1995) and reconcile with the claim that index fund inflows resemble a rightward shift in the demand curve resulting in a positive return shock.

As a robustness check, I regress index fund flows on lagged returns to confirm that the results of Table 4b cannot be explained by investors trying to chase returns, at least at a monthly time horizon. The results of these regressions are found in Table 4c. In Regressions (1) and (2), I find no significant relationship between returns from periods $t-1$ to $t-3$ with current scaled index fund flows, indicating that index fund flows do not display return-chasing behavior. Although I cannot reject the hypothesis that this return-chasing behavior does not occur on a shorter daily or weekly timescale due to data limitations, it is unlikely that such behavior would disappear over the longer horizon. In Regressions (3) and (4), I run the same tests on the unexpected index fund flows and find similar results, indicating no significant return-chasing behavior in either measure of S&P 500 passive flows.

In summary, I find a statistically significant relationship in the short run between passive S&P 500 fund flows and underlying index return. This indicates a short-run rightward shift in demand from passive funds during periods of high inflows and a corresponding increase in the prices of index components, providing some support to Shleifer's (1986) interpretation of index inclusion effects. Passive

S&P 500 fund flows do not exhibit return chasing behavior and the flows to individual passive index funds appear to be relatively uncorrelated.

3.2 Lagged Passive Flows and Index Returns

To analyze longer term negative correlations between fund flows and future index returns, I conduct lagged regressions of past pooled flows on contemporary returns over various time horizons, including $t-12$, $t-9$, $t-6$, and $t-3$. Ackman's theory would predict that these regressions would show a negative effect of passive fund flows over the longer horizon as the long term rate of return decreases for overvalued index components buoyed by short-term price pressures. I conduct these regressions using both geometrically-weighted and equal-weighted past flows. Equal-weighted flows equally weight flows from each previous time step, while geometrically-weighted regressions weight more recent flows more than older flows using geometric discounting. Mathematically, the two sets of regressions can be summarized as the the following:

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=t-n}^{t} F_i, \quad n \in (3, 6, 9, 12) \tag{11}$$

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=t-n}^{t} \sigma^i \times F_i, \quad n \in (3, 6, 9, 12) \tag{12}$$

The parameter values of σ vary from 0 to 1, much like many models of extrapolation in the literature, including Barberis, Greenwood, Jin, and Shleifer (2015). Functionally, σ serves as a discount factor for past flows.

I document a significant negative relationship between lagged passive S&P 500 fund flows and contemporaneous return, indicating the short term positive price impact reverting over a longer horizon. Table 5 documents this phenomenon, where pooled lagged index fund flows result in negative contemporary returns. This result holds after controlling for flows to actively managed funds, as documented in Regression (2). Interestingly, I observe no impact of actively managed fund flows in either direction. Regressions (3) and (4) document similar effects when discounting flows geometrically (providing more weight to more recent flows). The negative relationship between pooled lagged flows and contemporaneous return supports the axiomatic efficient markets belief that friction-based movements which may create short-term distortions are corrected in the longer term as prices reflect fundamental value.

The values of the regression coefficients in this set of regressions indicate the change in contemporaneous S&P 500 return for each one unit increase in scaled flows. Regression (1) indicates that a one unit increase in Pooled Scaled Flows (previous six months of pooled flows into S&P 500 tracking flows with a cumulative value of the 1% of total S&P 500 market cap) results in a 5.6% drop in monthly return. While the Adjusted R^2 of the regression is only 0.022, this again is not surprising, as one would expect significantly more drivers of S&P 500 returns than solely past fund flows.

3.3 Model of Flow-Price Relationship

Both of these results (short-term positive relationship between flows and returns coupled with longer-term reversion) could be generated by assuming that index levels follow a modified AR(1) process with a flow component and an upwards trend[12].

$$P_t = \gamma F_t + \phi P_{t-1} + \beta t + \epsilon_t \tag{13}$$

If I assume that the impact of flows to price effects is linear to the amount of flows (as previous OLS regressions have), a one period flow shock will result in a short term increase in price (representing the first result of short term outperformance). Due to the mean reverting nature of the AR(1) process, for each subsequent period, this flow

12. Here, P_t indicates price of the index, P_{t-1} previous period index price, F_t scaled flows over the period , ϕ a parameter for autoregressive behavior, β a parameter for time dependence, ϵ_t a random variable.

would have moderately negative impacts on returns in future periods based on the value of ϕ, assuming that it is less than one. Assuming some persistence of higher than normal flows, the model allows for persistently higher than normal returns over a longer time horizon, yet this will correct as flows subside. Bt allows for an upwards drift over time, consistent with historical market movements. This model allows for explanation of both of these results. Fitting this model to the data I have on the S&P 500, I find significance in all of these variables at at least the 90th percentile and an adjusted R^2 of .9914. I estimate the value of γ as 116.662 with a t-value of 2.107, the value of β as 0.100 with a t-value of 1.776, and the value of ϕ as .981 with a t-value of 89.234.

3.4 Passive Flows and Relative Mispricing

Having established a relationship between S&P 500 Index Fund & ETF flows and underlying index returns and provided some theoretical framework, a next step involves understanding whether these flows disproportionally affect the performance of stocks within the S&P 500 relative to similar companies outside the index. As a first pass, I evaluate the relationship between the spread between S&P 500 and S&P 400 returns and S&P 500 passive fund flows. Each period, I construct a long-short portfolio long the S&P 500 and short the S&P 400 in equal weights. The S&P 400 Mid Cap Index tracks the next 400 stocks with market capitalizations under the inclusion cutoff for the S&P 500. It thus provides a meaningful point of comparison to determine whether S&P 500 index fund flows only effect index constituents or extend to similar companies outside of the index. However, this comparison is not perfect, as companies within the S&P 500 may not directly be comparable to companies within the S&P 400, especially at the extremes of the two indexes (e.g. comparing the largest company in the S&P 500, Apple, with a $735.2B market capitalization to the smallest company in the the S&P 400, Vista Outdoor with a \$1.12B market capitalization)[13].

I regress the return of the S&P 500/S&P 400 long-short portfolio on scaled flows, adjusted flows (as defined above), previous period portfolio return (as a robustness check), and active flow controls. Mathematically, these regressions can be summarized as the following:

$$R_{SP500,t} - R_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1 F_t + \hat{\beta}_2(R_{SP500,t-1} - R_{SP500,t-1}) \tag{14}$$

$$R_{SP500,t} - R_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6}) + \hat{\beta}_2(R_{SP500,t-1} - R_{SP500,t-1}) \tag{15}$$

$$R_{SP500,t} - R_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1 F_t + \hat{\beta}_2 ActiveFlow_t + \hat{\beta}_3(R_{SP500,t-1} - R_{SP500,t-1}) \tag{16}$$

$$\begin{aligned} R_{SP500,t} - R_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6}) + \hat{\beta}_2 (ActiveFlow_t - \sum_{j=t-6}^{t} \frac{ActiveFlow_j}{6}) \\ + \hat{\beta}_3(R_{SP500,t-1} - R_{SP500,t-1}) \end{aligned} \tag{17}$$

The first specification asks whether the S&P 500-S&P 400 return spread depends on scaled S&P 500 passive flows, controlling for previous period return spread. The second specification uses the adjusted unexpected flows metric. The third and fourth specifications include active fund flow controls. I report the results of these regressions in Table 6. I document no significant relationship between scaled S&P 500 ETF and index fund flows and the return of this portfolio, regardless of controls used. I find a statistically significant negative relationship between scaled active flows

13. Market capitalization data as of March 21, 2017.

and the return of this portfolio. This could be indicative of active managers more frequently targeting stocks outside of the S&P 500 due to perceived overvaluation, or statistical noise. I find no significant relationship between the previous return spread and the current period spread. The results of regressions of the return spread of the S&P 500 and S&P 400 on passive flows indicate no strong impact on the relative pricing of companies within the S&P 500 versus out-of-index peers driven by S&P 500 index-tracking instruments.

To more narrowly identify the specific impact of flows on relative stock performance of similar securities, I construct a portfolio consisting of a long position in the bottom twenty stocks of the S&P 500 by market capitalization and a short position in the top twenty stocks of the S&P 400 by market capitalization, updated every month. I use this portfolio to control for the disparity in company characteristics between the largest companies within the S&P 500 and the smallest companies with the S&P 400 and to isolate the effects of flows right at the border between the two indexes. Restricting the long-short portfolio to these 40 stocks dramatically restricts differences in market capitalization. In fact, the average market capitalization for the top twenty S&P 400 companies exceeds the average market capitalization for the bottom twenty S&P 500 companies in 65% of months in the sample. Following the same method as I use to construct the S&P 500/S&P 400 portfolio above, I create a portfolio long \$1 of the bottom 20 companies of the S&P 500 and short \$1 the top 20 companies of the S&P 400. While this method reduces the impact of size effects, it does limit diversification by industry, in that there are months of the sample where the companies in the portfolio are particularly concentrated in one sector, such as financials or technology. To mitigate these effects, in all regressions involving this portfolio I control for industry fixed effects[14].

14. I create indicator variables for each Global Industry Classification System Sector (GICS Sector). There are 10 major sectors; I create an indicator variable for each sector and record the count of each sector in each portfolio and include these variables as controls in my regressions.

Much like the previous return differential tests above, I regress the performance of this portfolio on scaled S&P 500 passive flows, controls, and unexpected flows. Mathematically, these regressions are identical to those in Equations (14), (15), (16), and (17) above. Similarly to the general S&P 500 versus S&P 400 case, and perhaps surprisingly, I find no relationship between flows and the return of the border portfolio. My regression results can be found in Table 7 below. This result indicates no disproportionate price impact to the S&P 500 from S&P 500 passive fund flows versus out-of-index peers.

While I find no evidence of relative mispricing when looking at contemporaneous passive fund flows and portfolio returns, perhaps mispricing may occur on a longer time scale and be undetectable in monthly regressions. Beyond return spreads, distortion effects could manifest themselves in differences in relative valuation. Two common valuation metrics used by investors and economists alike include the book-market ratio and the the price-earnings ratio.

The book-market ratio, often expressed in the inverse as the price-book ratio by practitioners, is the ratio of the book value of a company's equity on their balance sheet (book value of assets - book value of liabilities) to the market value of equity (market capitalization). Companies with higher book-market ratios (lower price-book ratios) than matched peers are perceived to be "cheaper" than their peers and are considered "value" stocks. Another common valuation ratio used by practitioners especially is the price-earnings ratio. This ratio divides the market capitalization of a company by the company's post-tax net income. Which metric of earnings used often varies, but most commonly is last twelve months (LTM) net income, analyst estimates of next twelve months (NTM or Forward) net income, or some mix of the two. In calculating price-earnings ratios in this paper, I use LTM net income as the denominator of the price-earnings calculation. Companies with lower price-earnings ratios are considered "cheaper" than their peers as investors have to pay less for a given dollar of earnings. Evaluating the impact of S&P 500 passive fund flows on these valuation metrics for stocks within the S&P 500 and out-of-sample peers allows for a longer term test of the impact of flows on relative valuation.

In order to test for relationships between flows into index-linked products and the relative valuations of index components, I additionally require data on stock specific fundamentals such as earnings per share (EPS) and book value in order to construct book-market and price-earnings ratios. For these metrics, I use the CRSP/Compustat Merged database, which merges identifiers in the CRSP security database with those in the Compustat database. Compustat provides information on company fundamentals, including financial ratios, operating metrics, and identifying characteristics for over 99,000 securities, comprising over 99% of global market capitalization. Compustat provides quarterly data from 1962 until present and is linked to CRSP pricing data.

For each of these two metrics, I create portfolios mirroring those I create above for the S&P 500 versus the S&P 400 and the bottom twenty stocks of the S&P 500 versus the top twenty stocks of the S&P 400. Instead of using value-weighted returns, I use the calculated book-market or price-earnings ratio. I subselect only stocks with positive price-earnings ratios in either portfolio in order to avoid impacts of companies with negative earnings. More precisely, a company with very slightly negative earnings will register as having a very negative price-earnings ratio, skewing the portfolio's price-earnings ratio dramatically. I additionally exclude any companies with a price-earnings ratio exceeding 100, for similar distortive reasons.

Figure 5 below shows both the S&P 500 and S&P 400 price-earnings ratios over time. I regress the S&P 500 - S&P 400 price-earnings ratio spread on cumulative past flows. To account for significant autocorrelation in observations of price-earnings ratio, I include the previous spread as an explanatory variable.

Mathematically, for price-earnings ratios, these regressions take the following form:

$$PE_{SP500,t} - PE_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP500,t-1} - PE_{SP400,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} F_i \qquad n \in (3,6,9,12) \tag{18}$$

$$PE_{SP500,t} - PE_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP500,t-1} - PE_{SP400,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} F_i + \hat{\beta}_3 \sum_{j=t-6}^{t} ActiveFlow_j \qquad n \in (3,6,9,12) \tag{19}$$

$$PE_{SP500,t} - PE_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP500,t-1} - PE_{SP400,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} (F_i - \sum_{j=i-6}^{j=i} \frac{F_j}{6}) \qquad n \in (3,6,9,12) \tag{20}$$

$$\begin{aligned} PE_{SP500,t} - PE_{SP400,t} &= \hat{\beta}_0 + \hat{\beta}_1(PE_{SP500,t-1} - PE_{SP400,t-1}) \\ &+ \hat{\beta}_2 \sum_{i=t-n}^{t} (F_i - \sum_{j=i-6}^{j=i} \frac{F_j}{6}) \\ &+ \hat{\beta}_3 \sum_{i=t-n}^{t} (ActiveFlow_i - \sum_{j=i-6}^{j=i} \frac{ActiveFlow_j}{6}) \quad n \in (3,6,9,12) \end{aligned} \tag{21}$$

The first specification examines the relationship between S&P 500 - S&P 400 price-earnings spreads and pooled S&P 500 passive tracking fund flows. The third test uses adjusted unexpected flows, and the second and fourth specifications add controls for actively-managed fund flows. All four regressions control for the previous period price-earnings spread. I construct similar portfolios for the bottom twenty stocks of the S&P 500 versus the top twenty stocks of the S&P 400. In the interest of space, I have not reproduced them here. The results of these regressions can be found in

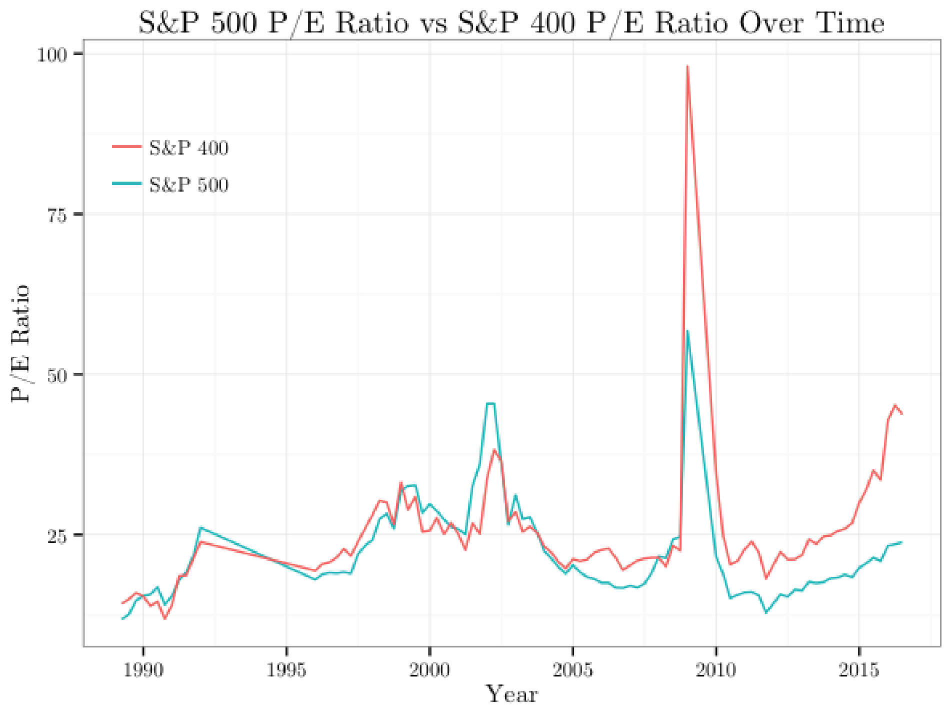

Figure 5: S&P 500 and S&P 400 P/E Ratios have mirrored one other over the entire sample. To construct this sample, I use data on index price and earnings-per-share from the Compustat US Database. All negative observations and observations exceeding 100 have been removed (May 2009).

Tables 8 and 9. I run regressions with the same specifications using the book-market ratio and report the results in Table 10. I find no significant relationship between the spread of the S&P 500 versus S&P 400 price-earnings ratio and S&P 500 index fund or ETF flows. When I narrow my sample to the bottom twenty companies in the S&P 500 by book value and the top twenty companies in the S&P 400 by book value, this result holds. For book-market ratios, I find no relationship between the spread of book-market ratios and S&P 500 index fund or ETF flows. While the lack of a significant result is perhaps surprising, it aligns with the previous tests of return differentials.

Concerning the impacts of the growth of indexing on price levels, I find that within the S&P 500, passive fund flows have a short term positive impact on price levels; in the long run, the effect reverts and large inflows indicate lower levels of future return. However, these effects do not result in any statistically significant relative mispricing of S&P 500 components versus similar out-of-sample peers within the S&P 400, even when isolating companies with similar market capitalization and controlling for industry fixed effects. I observe no relative mispricing when analyzing either data on stock returns or valuation ratios. While this result provides some support for a demand-based theory of short-term asset pricing, it does not demonstrate a significant distortion in the valuation of components in heavily-followed indexes such as the S&P 500 as a result of index fund and ETF demand.

4 Volatility & Comovement

Indexation also has an impact on the comovement of stocks with other stocks inside and outside of the index. Studies on comovement effects extend the effects of indexation beyond short term returns. A considerable body of literature demonstrates that once a security is added to an index, it begins to comove much more strongly to other securities within the index. Within the S&P 500, Barberis, Shleifer, and Wurgler (2005) document a nearly four-fold increase in security beta to the S&P 500 in the 24 months after index inclusion coupled with a 65% decrease in security beta to all non-S&P 500 securities. Greenwood and Sosner (2007) see similar effects in the Japanese Nikkei 225 index; Coakley and Kougoulis (2004) document even larger effects within the FTSE 100. In this paper, I replicate and update Barberis, Shleifer, and Wurgler (2005) tests and include more recent data. I then ask whether the impact of index investing on comovement has been increasing over time in line with inflows into S&P 500 tracking funds.

Theories of stock comovement generally are divided into three major models: a more traditional, fundamentals-based view; a category-based view; a habitat-based view[15]. The fundamentals-based view of return comovement asserts that securities comove with securities with similar fundamental characteristics and underlying exposures (such as industry, geography, supply chain, and common customers). The category-based view argues that investors group stocks into categories such as "value stocks", "large stocks", and "growth stocks" and adjust their portfolios to have certain exposures to each category. Since individual stocks in each category effectively are perceived as identical to these investors, whenever investors shift exposure in and out of the category, all stocks within the category move accordingly. The habitat-based view postulates that investors restrict their trading to certain "habitats" of stocks such as U.S. stocks, S&P 500 stocks, and that stocks in similar habitats thus move together. If an inclusion or deletion to/from an index results in a significant change in the comovement patterns of a stock, it provides evidence towards the habitat-based and category-based views of comovement. In this section I evaluate the relationship between the growth of index funds and ETFs and the changes in comovement patterns of S&P 500 stocks following an inclusion or exclusion event.

I begin with the construction of a sample of inclusions and exclusions within the S&P 500 over the period from January 1989 to December 2016. I use the Compustat North America Index Constituents file to identify all S&P 500 index changes over the period. I then utilize the Compustat/CRSP Merged database to link index changes to returns data. I exclude all events without sufficient data on both sides of the event, meaning that all firms included in my sample must have 36 months of available return history both prior to and following the event date. This excludes firms added to or removed from the index due to M&A activity, corporate spin offs, privatization, and bankruptcy. These events characterize a large percentage of index deletions, and my final sample consists of 394 inclusions and 145 deletions.

15. Barberis, Shleifer, and Wurgler (2005) provides additional detail on these three categories.

4.1 Univariate Changes in Comovement

Following Barberis, Shleifer, and Wurgler (2005), I begin with univariate regressions to estimate the beta of individual stocks to the S&P 500 in the 36 months before and after the inclusion or deletion event to measure whether comovement with the index differs before and after the event. I use the *effective date* of the event as the breakpoint between the pre and post event regressions, unlike Barberis, Shleifer, and Wurgler (2005) who use the *announcement date* as the endpoint of the pre-event regressions. I choose to use this date exclusively since index funds and ETFs are constrained from investing in securities until formal addition to the index. Utilizing the effective date allows for isolation of index fund and ETF demand effects. For each inclusion and deletion event, I run the following regression:

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t} \tag{22}$$

where $R_{i,t}$ corresponds to the return of stock i in month t and $R_{SP500,t}$ corresponds to the return of the S&P 500 over the same month. For each event, I record $\Delta\beta_1$, the change in beta for the security with the S&P 500 before and after the event. I find a statistically significant change in beta for stocks added to the S&P 500 of .143, corresponding to a t-statistic of 3.395. This change appears to be declining after 2005. Figure 6a below shows the change in beta over time for inclusions, grouped by year. Prior to 2000, $\Delta\beta_1$ averaged 0.032. In the middle part of the sample, from 2000 to 2006, average $\Delta\beta_1$ increases to 0.302. In the final years of my sample, from 2007 to present, $\Delta\beta_1$ compresses to 0.071. In each year, there are at least 4 inclusions. Inclusions became more frequent over time, peaking in 2000 with 37 inclusions, before normalizing to approximately 15 inclusions per year. Table 11 includes a breakout of S&P 500 index changes by year.

However, for deletions, changes in beta to the S&P 500 follow the opposite direction as Barberis, Shleifer, and Wurgler (2005) predictions. Over the entire sample of deletions, I find a $\Delta\beta$ of .198, corresponding to a t-statistic of 2.1, indicating a statistically significant *increase* in comovement with the S&P 500. The magnitude of this increase exceeds that of the change in beta for inclusions (.143). There does not appear to be a significant change in magnitude of $\Delta\beta$ over time for deletions. Figure 6b below shows the change in beta over time for deletions (grouped by year). Importantly, compared to inclusions, there are comparatively less valid deletions per year from the S&P 500. In each year following 1989, there is at least one deletion, with an average of 6.04 over the sample. The lack of a significant number of valid index deletions every year results in difficulty in gauging annual trends, as individual outliers could skew results dramatically.

4.2 Bivariate Changes in Comovement

In addition to the univariate regressions, I additionally run bivariate regressions in the same vein as Barberis, Shleifer, and Wurgler (2005). These regressions test the individual security's correlation with both the S&P 500 and with the rest of the market. These regressions take the form:

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t} + \hat{\beta}_2 R_{exSP500,t} \tag{23}$$

To calculate the rest-of-market return ($R_{exSP500,t}$), I use the following identity:

$$R_{VWCRSP,t} = \frac{MKT_{CRSP,t-1} - MKT_{SP500,t-1}}{MKT_{CRSP,t-1}} R_{exSP500,t} + \frac{MKT_{SP500,t-1}}{MKT_{CRSP,t-1}} R_{SP500,t} \tag{24}$$

where $R_{VCRSP,t}$ indicates the return on the value-weighted CRSP Index which summarizes the returns of all stocks traded on the NYSE, Nasdaq, and AMEX, $MKT_{CRSP,t-1}$ indicates the total market capitalization of the value-weighted CRSP index, and

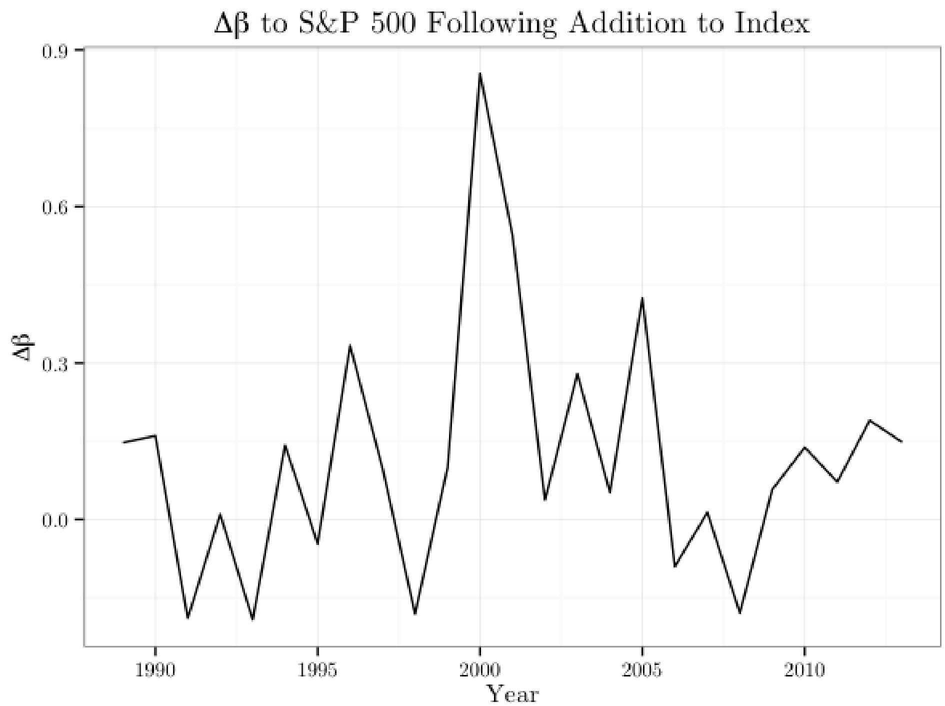

Figure 6a: Univariate changes in beta for inclusions have compressed in recent years but still appear economically significant. In this sample, there are a total of 394 inclusion events. To construct the sample, I measure changes of β_1 in the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t}$$

between the 36 months prior to inclusion and the 36 months immediately following.

$MKT_{SP500,t-1}$ indicates the total market capitalization of the S&P 500 index. Effectively I back into the rest-of-market return by subtracting the market capitalization-weighted contribution of the S&P 500 to total market return.

In my bivariate specifications, I find a similar result to Barberis, Shleifer, and Wurgler (2005): statistically significant increases in S&P 500 betas and decreases in rest-of-market betas for inclusions. I observe more significance both in terms of magnitude and statistical significance in the bivariate regressions, much like Barberis, Shleifer, and Wurgler (2005). The change in beta to the S&P 500 ($\Delta\beta_{SP500}$) averages a .404 increase upon inclusion, corresponding to a t-statistic of 4.78, the change in beta to the rest of the equities market ($\Delta\beta_{exSP500}$) averages a .265 decrease, corresponding

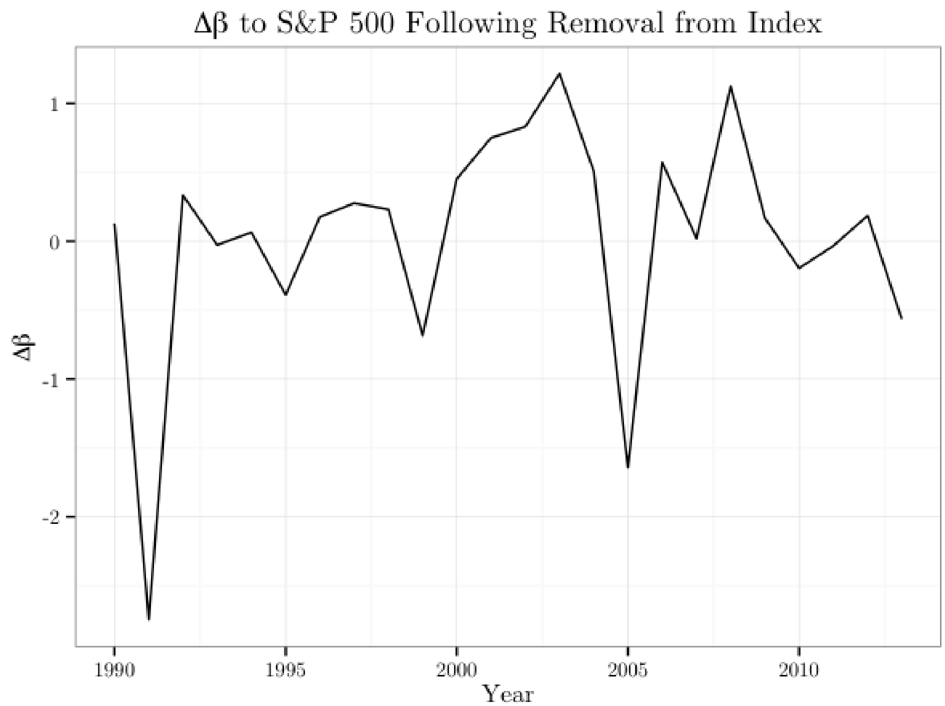

Figure 6b: Univariate changes in beta for deletions appear to be more volatile and potentially be in the opposite direction as Barberis, Shleifer, and Wurgler (2005) predictions. In this sample, there are a total of 145 deletion events. To construct the sample, I measure changes of β_1 in the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t}$$

over the 36 months prior to deletion and the 36 months immediately following.

to a t-statistic of -3.90. Figure 7a below shows the bivariate change in betas over time for inclusions (grouped by year).

For deletions, the direction of the effect mirrors Barberis, Shleifer, and Wurgler (2005), with decreases in S&P 500 betas and increases in rest of market betas. This contrasts with the results of the univariate regression, but changes in beta are less statistically significant. Deletions on average experience a .438 decrease in beta to the S&P 500 ($\Delta\beta_{SP500}$), corresponding to a t-statistic of -1.80, insufficient to reject the null hypothesis of no change in betas at a 95% significance threshold. On average, deletions experience a .640 increase in beta to the rest of the market ($\Delta\beta_{exSP500}$),

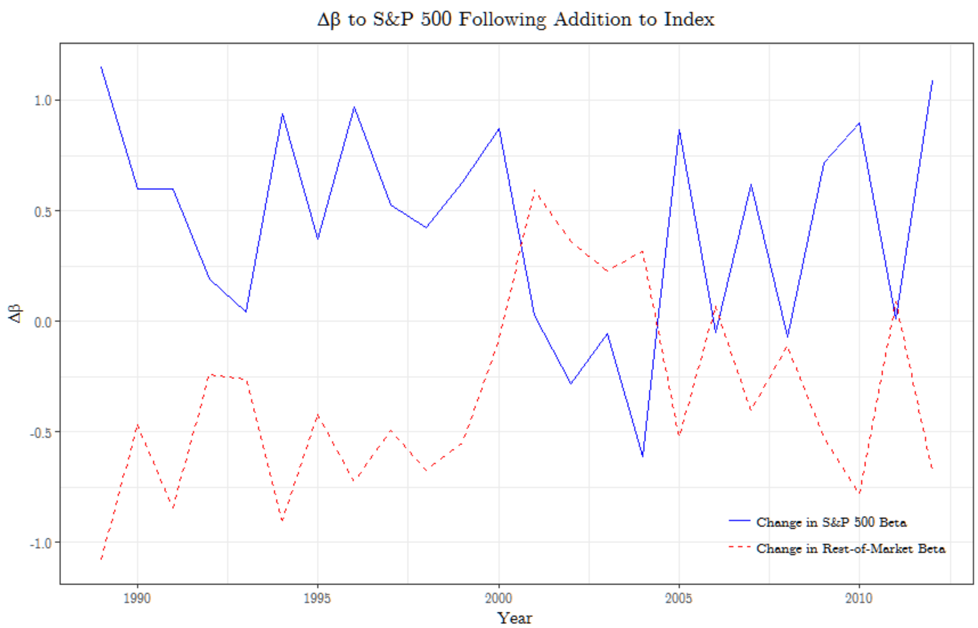

Figure 7a: Bivariate changes in beta for inclusions have compressed in recent years but still appear economically significant. In this sample, there are a total of 394 addition events. To construct the sample, I measure changes of β_{SP500} and $\beta_{exSP500}$ in the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t} + \hat{\beta}_2 R_{exSP500,t}$$

between the 36 months prior to addition and the 36 months immediately following. To calculate the rest of market return, I use the identity

$$R_{VWCRSP,t} = \frac{MKT_{CRSP,t-1} - MKT_{SP500,t-1}}{MKT_{CRSP,t-1}} R_{exSP500,t} + \frac{MKT_{SP500,t-1}}{MKT_{CRSP,t-1}} R_{SP500,t}$$

corresponding to a t-statistic of 2.95. Figure 7b below shows the bivariate change in betas over time for deletions.

4.3 Impact of Passive Flows on Magnitude of Comovement Effects

Figures 6 and 7 indicate no secular increase in the impact of index membership on comovement. However, the Barberis, Shleifer, and Wurgler (2005) effect appears to have remained persistent over time. To more precisely identify the relationship between the growth of indexing and comovement, I test for the relationship between

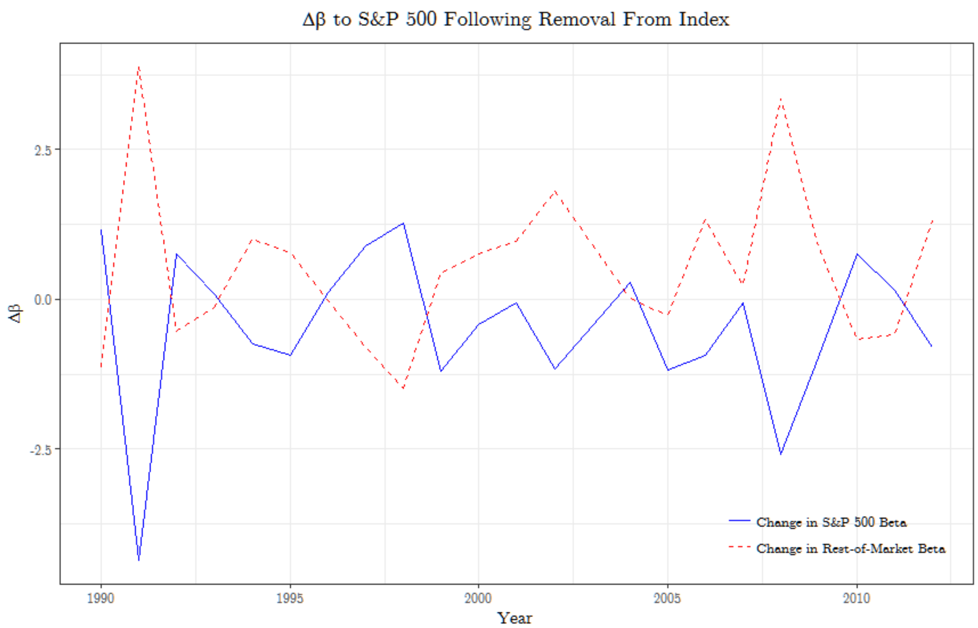

Figure 7b: Bivariate changes in beta for deletions appear very noisy, but follow Barberis, Shleifer, and Wurgler (2005) predictions. In this sample, there are a total of 145 deletion events. To construct the sample, I measure changes of β_{SP500} and $\beta_{exSP500}$ in the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t} + \hat{\beta}_2 R_{exSP500,t}$$

between the 36 months prior to deletion and the 36 months immediately following. To calculate the rest of market return, I use the identity

$$R_{VWCRSP,t} = \frac{MKT_{CRSP,t-1} - MKT_{SP500,t-1}}{MKT_{CRSP,t-1}} R_{exSP500,t} + \frac{MKT_{SP500,t-1}}{MKT_{CRSP,t-1}} R_{SP500,t}$$

flows and the change in betas for each event. I regress the change in security beta on cumulative scaled flows over the 72-month event period, the 12 months immediately surrounding the effective date of inclusion/deletion, the share of the S&P 500 owned by index funds and ETFs at the effective date, and the standard deviation of flows over the 72 month and 12 month periods around the effective date. These regressions take the following form:

$$\Delta\beta_{SP500,i} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \epsilon_i, \quad n \in (6, 36) \tag{25}$$

$$\Delta\beta_{SP500,i} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \hat{\beta}_2 \sum_{i=t-n}^{t+n} ActiveFlow_i + \epsilon_i, \quad n \in (6, 36) \tag{26}$$

$$\Delta\beta_{SP500,i} = \hat{\beta}_1 \hat{\sigma}_t + \epsilon_i = \hat{\beta}_1 \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2} + \epsilon_i, \quad n \in (6, 36) \tag{27}$$

$$\Delta\beta_{SP500,i} = \hat{\beta}_1 \sum_f \frac{NAV_{f,t}}{IndexCap_t} \tag{28}$$

I construct mechanically identical analogues of these equations for the changes in security betas to the rest of the market ($\Delta\beta_{exSP500}$). I repeat these regressions for S&P inclusions and deletions. The results of these regressions for inclusions and deletions can be found in Tables 12 and 13 below. I find no significant relationship between the magnitude of these changes in beta and any explanatory variables that I have selected, with the exception of the standard deviation metric for the change in multivariate betas for exits.

Economically, the strong impact of the standard deviation metric of flows on changes in multivariate betas for exits could be explained by the fact that since many of the companies removed from the S&P 500 for reasons unrelated to mergers and acquisitions activity are removed for negative share performance, index funds and ETFs may constitute a larger percentage of ownership in exits. Active managers may hold less of these securities given previous underperformance due to behavioral factors such as overreaction or perceived fundamental weakness. This results in a large component of trading volume (and thus price movements) coming from index investors, which then disappears upon removal from the index. When flows are volatile, more variation in the underlying price comes from these index buyers. This effect would not be as strong in index inclusions, which typically experience strong outperformance prior to being added to the index, as noted by Chen, Singal, and Whitelaw (2016). As a result of this outperformance, more active managers may notice the stock and contribute to its trading volume and price movements prior to its addition to the index.

Overall, while I find one intriguing result on the impact of the volatility of S&P 500 tracking fund flows on comovement in S&P 500 exits, on the whole I find limited evidence of a strong relationship between the growth of index investing and comovement effects if stocks entering and exiting the index.

5 Price Informativeness & Liquidity

Index membership may also have an impact on the liquidity of securities within the index. Liquidity measures the difficulty of trading a particular security; a security is considered liquid if agents can easily trade in and out of a security bearing minimal costs and illiquid if agents must accept significant financial and time costs in order to trade the security. Theoretically, risk-averse investors must receive some compensation for holding illiquid securities and this hidden cost must be included in asset pricing. A considerable body of economic literature concerns itself with the market microstructures that originate illiquidity, such as the bid-ask spread and constrained investors[16]. Another segment of previous work connects changes in liquidity to cross-sectional asset returns. Amihud, Mendelson, and Pederson (2006) demonstrate how adding liquidity measures to traditional asset pricing models helps explain cross-sectional stock returns, treasury yield curve arbitrage strategies, corporate bond yield spreads, and comparative pricing of claims to identical cash flows (such as illiquid derivatives versus underlying stocks). Amihud (2002) finds a significant positive relationship between his measure of illiquidity and cross-sectional stock returns.

The foundation of liquidity-based asset pricing can largely be linked to Grossman and Stiglitz (1980). Grossman and Stiglitz argue against a perfectly

16. See Brennan and Subrahmanyam (1996), O'Hara (1995), Harris (2003), Madhavan (2000), and others.

information-efficient market and coined the term now known as the "Grossman Stiglitz Paradox", that for markets to be perfectly efficient, arbitrageurs must arbitrage away any inefficiencies and for arbitrageurs to exist, sufficient compensation must accrue to them to offset the cost of obtaining information (salaries, rent, data providers). The paradox exists because in perfectly efficient markets where prices reflect public information arbitrageurs would not be compensated for obtaining information. Thus, perfectly efficient markets cannot exist. This implies that some frictions must exist even in equilibrium to allow arbitrageurs to survive, including liquidity frictions.

With most measures of liquidity, it seems likely that the addition of a stock to a highly-traded and followed index such as the S&P 500 would result in significant increases in liquidity. This follows from the large amount of investors who invest solely in the S&P 500 as well as index funds and ETFs which must track the S&P 500. Using the same data set that I use to calculate changes in beta to the S&P 500 upon index inclusions and exclusions, I calculate changes in liquidity measures and the relationship of these changes to the level of index fund and ETF flows over the period.

5.1 Selecting a Liquidity Measure

To test the impact of S&P 500 membership on stock liquidity, I first choose which liquidity measure to use in my analysis. Conventional measures of liquidity can be categorized into three major categories: transaction cost measures, volume-based measures, and market impact measures[17]. Transaction costs include implicit execution costs and explicit order processing and tax costs; the most common metric of transaction costs used is the bid-ask spread, which captures both of these costs. Volume-based measures seek to quantify liquidity in terms of the depth of a market and quantity of trading activity. The most common volume metrics of liquidity used

17. For a more in-depth discussion of these four major categories, see Sarr and Lybek (2002).

are trading volume and share turnover (which adjusts trading volume for the total number of shares outstanding). Market impact measures seek to explain the residuals of traditional asset pricing models (e.g. CAPM) by trading volume or bid ask spread changes. These measures consider illiquidity to be strong price responses to order flow. The most common market impact measure used is Kyle's lambda, based off of Kyle (1985), which mathematically is represented as $\frac{|\Delta Price_t|}{Volume_t}$. Kyle's lambda is typically measured over very short time horizons[18] and seeks to isolate the impacts of trading on prices. On a broader scale, market impact measures connect the scale of price impact to trading frequency.

I use a modified version of Amihud's (2002) measure of liquidity as my primary liquidity measure in all empirical tests. Similar to Kyle's lambda, Amihud's (2002) measure measures the impact of trading volume on prices, except averaged over a longer horizon. Within the finance literature, this metric has become one of the most commonly used estimators of security liquidity. As noted by Lou and Shu (2014), over the period from 2009-2013, the metric appeared in over one hundred papers published in the Journal of Finance, the Journal of Financial Economics, and the Review of Financial Studies. I choose to utilize Amihud's measure predominantly due to its well-established relationship with cross-sectional stock returns and due to its relative ease of construction with the data I have available. I modify Amihud's measure to be calculated on a monthly instead of an annual basis. This modification has been documented in the literature by Acharya and Pederson (2005). I use the following measure:

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

where where $ILL_{i,m}$ indicates the illiqudity (modified Amihud's) measure for security i in month m, $D_{i,m}$ indicates the number of trading days for security i in month m,

18. In Kyle (1985) Volume indicates the total value of trades in the period ($P * Q$), not the total number of shares traded (Q).

$r_{i,t}$ and $Vol_{i,t}$ indicates the daily return and volatility (respectively) of security i on day t. For S&P 500 inclusions and exclusions, I first test for changes in the average illiquidity measure before and after index inclusion or deletion. I then regress these changes in the illiquidity measure on index fund and ETF flows over the period.

5.2 Passive Flows and Level of Illiquidity

I begin by testing whether addition to or removal from the S&P 500 results in a significant change in illiquidity. Using the same sample of 394 index inclusions and 145 deletions I use for my comovement tests (described above in Section 4), I measure the average $ILL_{i,m}$ for each security in the 36 months prior to and after the event (either inclusion or exclusion) and take the difference as ΔILL_i for each event. I find a statistically significant average decrease in illiquidity (ΔILL) for stocks added to the index of -.0011, corresponding to a t-statistic of -2.049. While the magnitude of this change in illiquidity appears low, these stocks are already amongst the largest, most liquid public stocks. This change represents a 57.2% decrease in average ILL_i upon index inclusion. However, this figure is skewed based on a few outliers with increases in illiquidity exceeding 500%; using the median ΔILL of -0.00039, a more reasonable 20.1% median decrease in illiquidity occurs upon index inclusion. Figure 8a, below, shows the absolute change in illiquidity for S&P 500 inclusions, grouped by year. I winsorize the data at the 2.5th and 97.5th percentiles. I find that the absolute change in illiquidity decreases over time, suggesting that the positive liquidity impact of S&P 500 inclusion has declined even as flows and index share have increased. However, although the absolute change in illiquidity upon inclusion has declined, the percentage change in illiquidity could have increased (provided that overall market levels of illiquidity have decreased over time). Figure 9a, below, shows the percentage change in illiquidity for S&P 500 inclusions, grouped by year. On a percentage basis, changes in illiquidity upon S&P 500 index inclusion do not appear to significantly change in magnitude over the length of the sample.

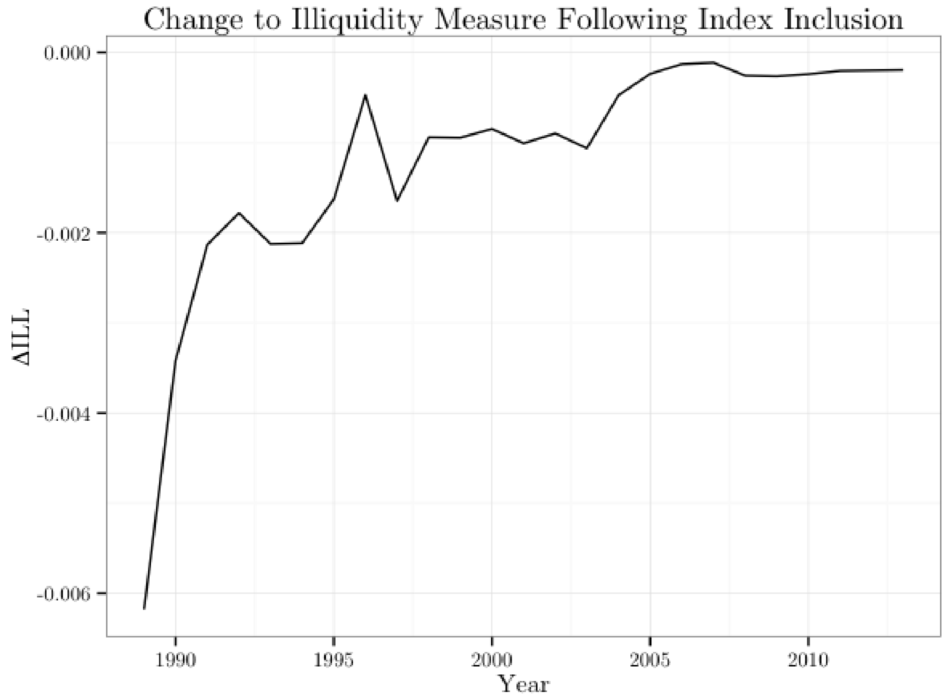

Figure 8a: Absolute changes in illiquidity measure following inclusion to the S&P 500 index have decreased over time, yet remain economically significant. In this sample, there are a total of 394 inclusion events. To construct the sample, I calculate the difference in average illiquidity over the 36 months prior to index inclusion and the the 36 months following index inclusion. I use Amihud's (2002) liquidity measure calculated on a monthly, rather than annual basis, and winsorize all data at the 2.5th and 97.5th percentiles.

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

For S&P 500 index deletions, I find a significant increase in absolute illiquidity. Over the total sample of deletions, I find a (ΔILL) of .0451, corresponding to a t-statistic of 2.321. Notably, prior to deletion, these stocks already have on average *six times* higher ILL_i measures than their peers within the S&P 500. Using the median ΔILL of 0.00521 for the same reasons as above (large outliers), deletion results in an median increase in illiquidity of 80.1%. Figure 8b, below, shows the absolute change in illiquidity over time for deletions, grouped by year. I winsorize

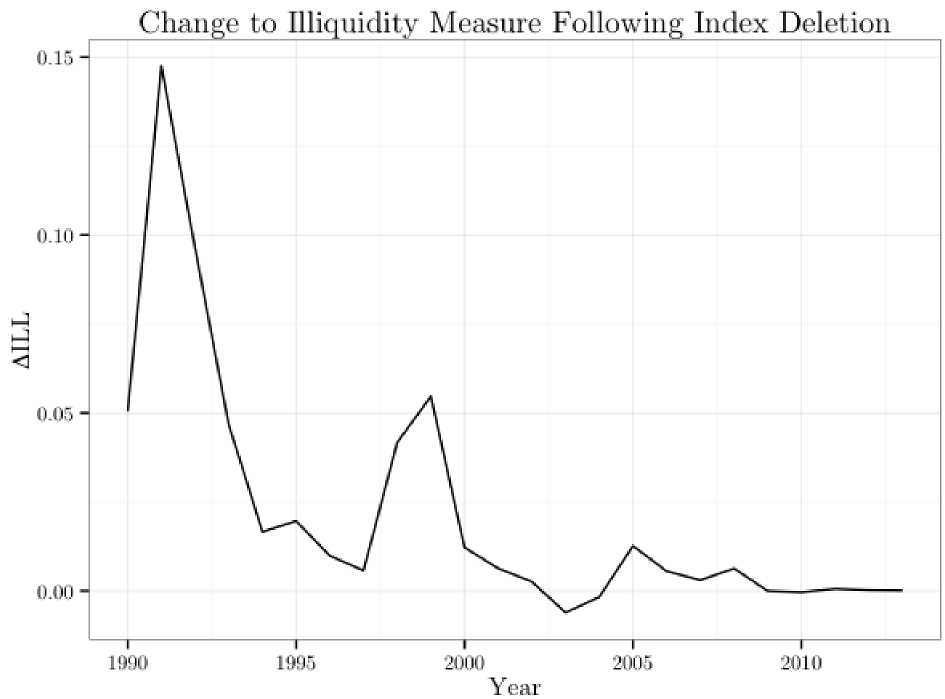

Figure 8b: Absolute changes in illiquidity measure following removal from the S&P 500 index have decreased over time, yet remain economically significant. In this sample, there are a total of 145 deletion events. To construct the sample, I calculate the difference in average illiquidity over the 36 months prior to index deletion and the the 36 months following index deletion. I use a Amihud's (2002) liquidity measure calculated on a monthly, rather than annual basis and winsorize all data at the 2.5th and 97.5th percentiles.

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

the data at the 2.5th and 97.5th percentiles. Much like in the inclusions case, I find that the absolute change in illiquidity decreases over time. Figure 9b shows the percentage change in illiquidity over time for deletions. Similarly to the case of S&P 500 inclusions, percentage changes in illiquidity on S&P 500 deletion do not appear to significantly change in magnitude over the length of the sample.

Figure 9 indicates no significant time-dependent trend in the impact of S&P 500 index inclusion and deletion on security liquidity. As expected, for index inclusions, illiquidity significantly decreases. The reverse holds true for deletions. To

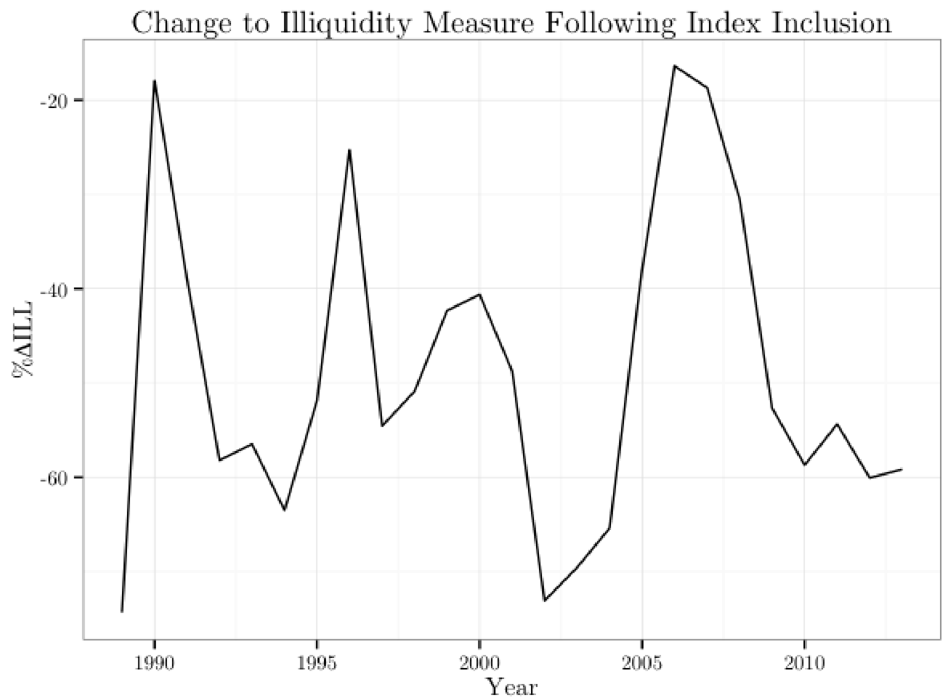

Figure 9a: Percentage changes in illiquidity measure following inclusion to the S&P 500 index do not appear to have grown significantly over time. In this sample, there are a total of 394 inclusion events. To construct the sample, I calculate the difference in average illiquidity over the 36 months prior to index inclusion and the the 36 months following index inclusion. I then divide this difference by the pre-event illiquidity in order to calculate percentage change. I use Amihud's (2002) liquidity measure calculated on a monthly, rather than annual basis.

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

formally examine the relationship between the growth of passive investing and changes in liquidity, I regress the percentage change in security illiquidity on the same factors I use in Section 4, including cumulative scaled flows into S&P 500 tracking vehicles over the 72-month event period and in the 12 months immediately following the effective date of inclusion/deletion, the share of the S&P 500 owned by index funds and ETFs at the effective date, and the standard deviation of flows over the 72 month and 12 month periods around the effective date. I choose to use the percentage change ($\%\Delta ILL_i$) instead of absolute change in order to separate the impact of passive funds

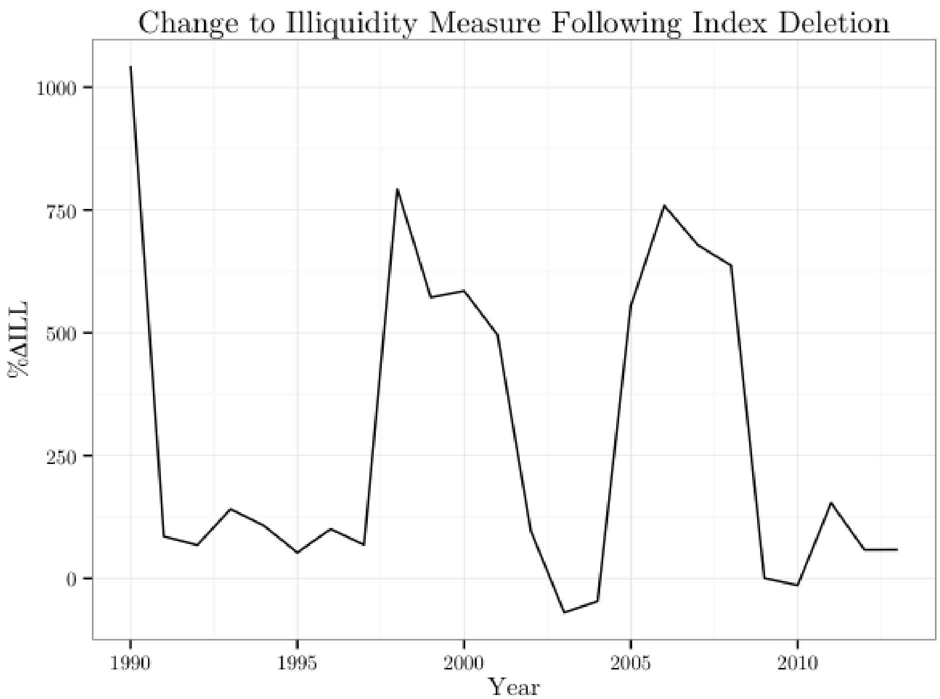

Figure 9b: Percentage changes in illiquidity measure following deletion to the S&P 500 index do not appear to have grown significantly over time. In this sample, there are a total of 145 deletion events. To construct the sample, I calculate the difference in average illiquidity over the 36 months prior to index deletion and the the 36 months following index deletion. I then divide this difference by the pre-event illiquidity in order to calculate percentage change. I use Amihud's (2002) liquidity measure calculated on a monthly, rather than annual basis.

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

from the market-wide general decrease in illiquidity. These regressions take the following form:

$$\%\Delta\overline{ILL_i} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \epsilon_i, \quad n \in (6, 36) \tag{29}$$

$$\%\Delta\overline{ILL_i} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \hat{\beta}_2 \sum_{i=t-n}^{t+n} ActiveFlow_i + \epsilon_i, \quad n \in (6, 36) \tag{30}$$

$$\%\Delta\overline{ILL_i} = \hat{\beta}_1\hat{\sigma}_t + \epsilon_i = \hat{\beta}_1\sqrt{\frac{1}{2n-1}\sum_{i=t-n}^{t+n}(F_i - \overline{F})^2} + \epsilon_i, \quad n \in (6, 36) \tag{31}$$

$$\%\Delta\overline{ILL_i} = \hat{\beta}_1\sum_f \frac{NAV_{f,t}}{IndexCap_t} \tag{32}$$

The first specification tests the impact of cumulative passive S&P 500 flows over the event period on percentage changes in illiquidity. The second specification adds a control for the active mutual fund flows over the period. The third specification tests for a relationship between the standard deviation of passive S&P 500 flows over the event period and percentage changes in illiquidity. The fourth specification tests for the relationship between passive share of the S&P 500 and percentage changes in illiquidity.

Table 14 (for inclusions) and Table 15 (for deletions) contain the results of these regressions. I find significant results in the opposite direction as one would expect. For inclusions, I document that increased passive S&P 500 flows actually result in a smaller increase in security liquidity upon index inclusion. For deletions, I identify no effects of increased flows to S&P 500-tracking index funds and ETFs on the magnitude of liquidity decreases for securities leaving the index. In all, the growth of passive investing in the S&P 500 does not indicate an increase the liquidity impact of index membership. In fact, larger passive fund flows seem to have narrowed the liquidity premium within the S&P 500, especially for index inclusions.

5.3 Passive Flows and Price Impact of Illiquidity

Additionally, I test to see whether beta loadings on the illiquidity measure change over the event period. To calculate loadings on the illiquidity measure, I use a univariate approach and run the following regression on either side of the event:

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 ILL_{i,t} \tag{33}$$

where $R_{i,t}$ corresponds to the return of stock i in month t and $ILL_{i,m}$ corresponds to the return of the illiquidity metric of stock i in month t. I record the change in $\hat{\beta}_1$ for each event and aggregate across all inclusions and exclusions in the set. Figure 10 below displays the changes in illiquidity loading over time for inclusions (10a) and deletions (10b). I find no statistically significant change in illiquidity loading for S&P 500 inclusions. For S&P 500 deletions, I find a significant increase in illiquidity loading of 17.23, corresponding to a t-statistic of 2.46. While the magnitude of this change appears high, ILL_i itself is very small. A one-unit absolute change in ILL_i represents a shift *25 times* the size of the observed S&P 500 index deletion effect on illiquidity. I observe an increased loading on illiquidity upon S&P 500 index deletion, but no significant change in illiquidity loading for S&P 500 inclusions.

I then run regressions to evaluate the effects of S&P 500 index fund and ETF flows on the security's change in loading on the liquidity factor. Given the lack of relationship between S&P 500 index inclusion and changes in illiquidity loading, I only run these regressions for S&P 500 deletions. The regressions take nearly identical forms to the regressions in Tables 14 and 15, except instead of ΔILL_i as the dependent variable, I use $\Delta\beta_{ILL}$. These regressions take the form

$$\Delta\beta_{ILL} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \epsilon_i, \quad n \in (6, 36) \tag{34}$$

$$\Delta\beta_{ILL} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \hat{\beta}_2 \sum_{i=t-n}^{t+n} ActiveFlow_i + \epsilon_i, \quad n \in (6, 36) \tag{35}$$

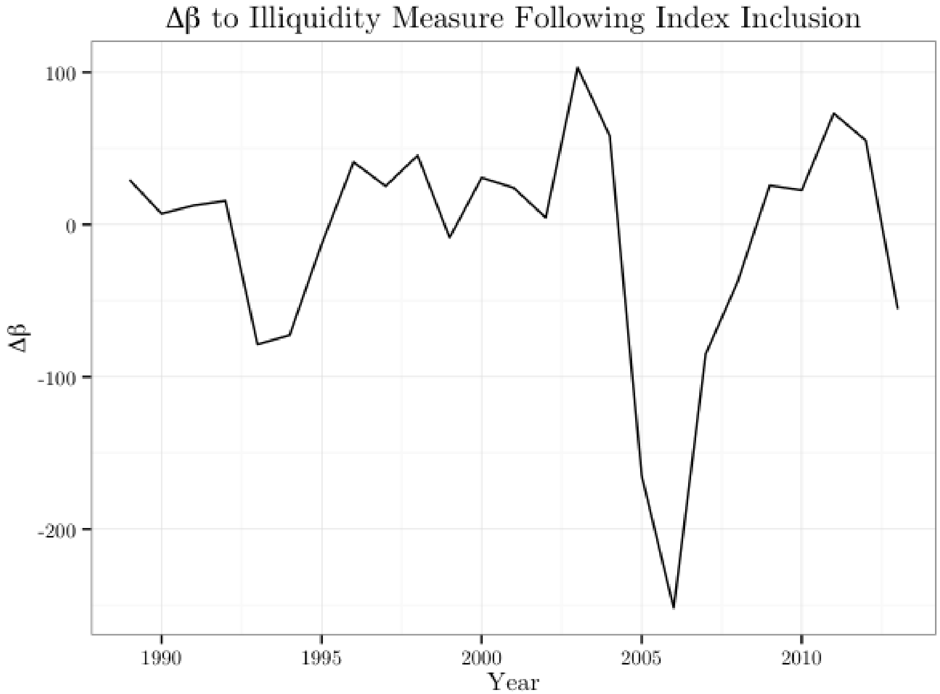

Figure 10a: Absolute changes in illiquidity loading following inclusion to the S&P 500 index have oscillated over time. In this sample, there are a total of 394 inclusion events. To construct the sample, I measure changes of β_1 in the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 ILL_i$$

over the 36 months prior to inclusion and the 36 months immediately following. I use Amihud's (2002) liquidity measure calculated on a monthly, rather than annual basis.

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

$$\Delta\beta_{ILL} = \hat{\beta}_1 \hat{\sigma}_t + \epsilon_i = \hat{\beta}_1 \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2} + \epsilon_i, \quad n \in (6, 36) \tag{36}$$

$$\Delta\beta_{ILL} = \hat{\beta}_1 \sum_f \frac{NAV_{f,t}}{IndexCap_t} \tag{37}$$

Table 16 contains the results of these regressions. I find that the illiquidity loading for S&P 500 index deletions has a positive relationship with passive fund flows and

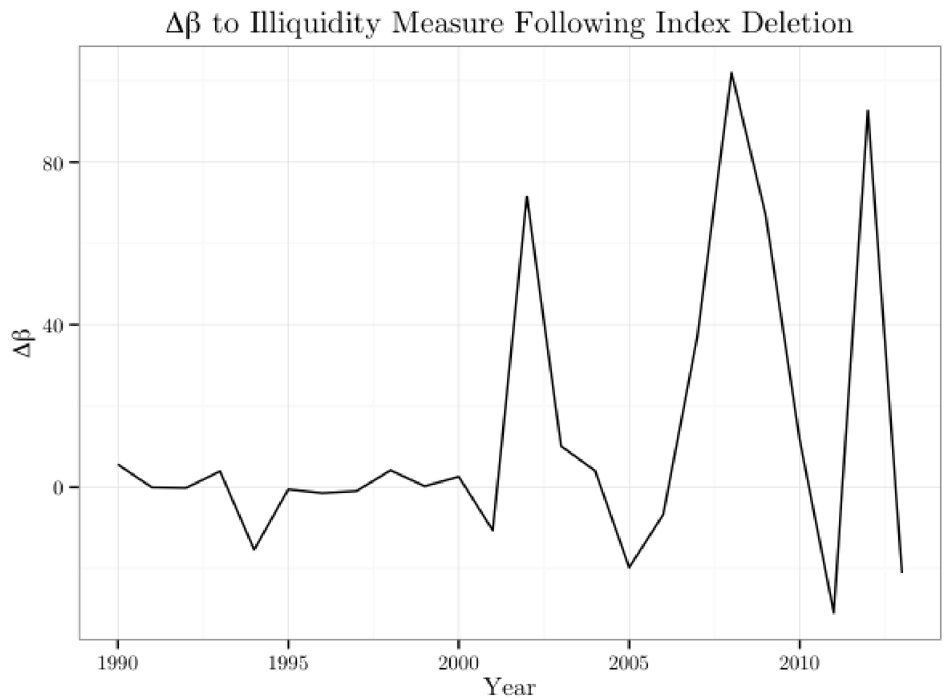

Figure 10b: Absolute changes in illiquidity loading following inclusion to the S&P 500 index have oscillated over time, but have been predominantly positive. In this sample, there are a total of 394 inclusion events. To construct the sample, I measure changes of β_1 in the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 ILL_i$$

over the 36 months prior to deletion and the 36 months immediately following. I use Amihud's (2002) liquidity measure calculated on a monthly, rather than annual basis.

$$ILL_{i,m} = \frac{1}{D_{i,m}} \sum_{t=1}^{D_{i,m}} \frac{|r_{i,t}|}{Vol_{i,t}}$$

the volatility of these flows. These results indicate that the S&P 500 deletion event results in security illiquidity becoming a more important predictor of return when index tracking flows are greater in magnitude or volatility. Economically, investors may perceive the liquidity drop upon index deletion as a larger concern when index membership (and the resultant passive flows) contributes a larger percentage of trading volume in the security. This could cause an increase in the premium investors need to hold the illiquid asset.

I find mixed evidence on the impact of the growth of passive management within the S&P 500 on the liquidity of index components upon addition or removal to/from the index. I identify relatively time-invariant decreases in illiquidity (median of 20.1%) for index inclusions and increases in illiquidity (median of 80.1%) for index deletions. These changes in fact narrow during periods of high passive flows and high passive flow volatility for index inclusions. I observe a statistically significant increase in illiquidity loading in index deletions. This loading increases in periods of high passive flows, indicating sensitivity to illiquidity increasing in periods of high passive S&P 500 fund flows.

6 Conclusion

I examine three common hypothesized effects of the growth in index investing on security characteristics from an empirical perspective. For the first effect, that index investing leads to changes in price level and relative distortion in pricing for stocks within and outside of an index, I find mixed evidence. I discover substantial evidence of a contemporaneous positive price pressure on index components as a result of index fund and ETF flows inflows into the S&P 500, and of longer-term market corrections for these flows. These findings support the interpretation of significant flows as a demand shock effecting prices in a modified AR(1) framework. However, I find no distortive effects when evaluating the relative valuation of securities within the S&P 500 versus close out-of-sample peers, either on a returns, P/E, or B/M basis.

A second well documented effect of indexing and passive investing concerns volatity and comovement. Barberis, Shleifer, and Wurgler (2005) document that upon addition to the S&P 500, stocks begin to comove much more strongly with the index and much less strongly with the rest of the market, with the reverse holding true for index deletions. I find that this change in betas has in fact compressed in recent years while S&P 500 passive fund flows have increased dramatically. Formally, I find no statistically significant relationship between the magnitude of changes in equity betas and S&P 500 passive fund flows.

A third potential effect of the growth of index investing involves the liquidity of securities within heavily followed indexes. I document a statistically significant increase in security liquidity upon S&P 500 index inclusion and a decrease upon deletion. However, the magnitude of the change in liquidity has decreased over time on an absolute basis. I do observe that—in the case of S&P 500 deletions—the sensitivity of contemporaneous stock return to liquidity increases with increased magnitude and volatility of flows.

My findings indicate the need for some temperance regarding perceptions of the impact of indexation and the growth of index investing. While I find some evidence of some of these three effects on security characteristics, I observe a limited connection between these effects and S&P 500 index fund and ETF flows, at least in monthly data. Perhaps the reason for these limited effects lies in the still limited size of index fund and ETF flows relative to total market capitalization, with the largest monthly shocks still less than $40 billion in an almost $20 trillion market. With increased availability of daily flow data for index funds and even intraday data for ETF creations and redemptions, perhaps these effects manifest themselves on a more granular scale and smooth out over the monthly horizon. Conceivably, these effects appear more frequently or to a larger degree in less liquid indexes than the S&P 500. Alternately, concerns regarding the size and scope of the economic effects of index investing on security characteristics may be largely overstated. While the seismic shift in the way that households and institutions manage their money from active to passive managers undoubtedly requires additional research and thought, my evidence indicates that at least as of now, practitioners and economists need not fear the effect of this trend.

7 Tables

Table 1: Full Sample Summary Statistics

Data on flows and returns sourced from CRSP. To identify S&P 500 passive vehicles I sample all funds in the CRSP Mutual Fund Names file flagged as ETFs or index funds and filter for all mentions of "S&P 500" or "500" within the fund_name field, and exclude all sector-specific, short-biased, or leveraged funds on an ad hoc basis. Scaled flows divide nominal flows by the market capitalization of the S&P 500 from the CRSP S&P Universe file. Adjusted Flows calculate the difference between current period flows and the six-month moving average of flows. S&P index returns are sourced from the CRSP S&P Universe file. Data on S&P inclusions and deletions comes from the Compustat North America Monthly file. Comovement βs are calculated as coefficients of the regression

$$R_{i,t} = \hat{\beta_0} + \hat{\beta_1} R_{SP500,t} + \hat{\beta_2} R_{exSP500,t}$$

in the bivariate case and

$$R_{i,t} = \hat{\beta_0} + \hat{\beta_1} R_{SP500,t}$$

in the univariate case. $\Delta\beta$s are calculated as the difference in the average βs from the 36 month period before the inclusion event and the 36 month period after the inclusion event. The same methodology is used in calculating illiquidity βs. I use a modified version of Amihud's (2002) illiquidity measure, calculated on a monthly basis.

Statistic	N	Mean	St. Dev.	Min	Max
Panel A: Flows and Returns (1989-2016)					
Flows and Indexed Assets					
Monthly Flows into Passive S&P 500 Funds ($MM)	333	1,041.229	6,157.638	−28,580.060	36,179.310
S&P 500 Passive Fund NAV ($B)	333	224,017.800	213,368.800	1,140.541	844,498.100
F_t	333	0.010	0.046	−0.207	0.222
Adjusted Flows (AF_t)	328	0.0002	0.043	−0.263	0.181
Scaled Active Fund Flows ($ActiveFlow_t$)	333	0.052	0.157	−0.588	0.450
Passive Fund Ownership of S&P 500 (%)	333	1.818	1.211	0.058	4.333
Returns					
S&P 500 Monthly Return (r_t)	333	0.007	0.042	−0.169	0.112
S&P 400 Monthly Return	333	0.010	0.048	−0.218	0.148
S&P 480-S&P 520 Portfolio Return	295	−0.053	0.083	−0.502	0.287
Panel B: Comovement $\Delta\beta$s					
Univariate Regressions: Inclusions					
$\Delta\beta_{SP500}$	394	0.142	0.831	−2.590	4.661
Univariate Regressions: Deletions					
$\Delta\beta_{SP500}$	145	0.198	1.130	−4.339	4.728
Bivariate Regressions: Inclusions					
$\Delta\beta_{SP500}$	394	0.433	1.671	−6.642	10.684
$\Delta\beta_{exSP500}$	394	−0.296	1.312	−6.128	4.439
Bivariate Regressions: Deletions					
$\Delta\beta_{SP500}$	145	−0.436	2.198	−11.740	4.489
$\Delta\beta_{exSP500}$	145	0.641	2.193	−5.648	11.163
Panel C: Liquidity Δ_{ILL}s and $\Delta\beta_{ILL}$s					
Inclusions					
Δ_{ILL}	394	−0.001	0.002	−0.007	0.001
$\%\Delta_{ILL}$	394	−21.2	379.8	−99.8	7307.1
$\Delta\beta_{ILL}$	394	−23.043	458.283	−2,871.760	5,602.480
Deletions					
Δ_{ILL}	145	0.045	0.234	−0.311	1.949
$\%\Delta_{ILL}$	145	483.1	1997.1	−82.0	18462.0
$\Delta\beta_{ILL}$	145	21.160	127.660	−386.601	963.083

Table 2: Correlation of Flows to Largest Three S&P 500 Tracking Funds

This table summarizes the correlation between flows into and out of the three largest S&P 500 tracking funds: SPY (SPDR S&P 500 ETF Trust), VFIAX (Vanguard 500 Index Fund Admiral Class), and IVV (iShares S&P 500 Index ETF). These three funds manage over $420B, or approximately 49% of total S&P 500 tracking assets. I find limited correlation between flows into each fund. I construct flows data by using fund net asset values and returns from the CRSP Mutual Funds file. Flows are scaled by total S&P 500 market capitalization from the CRSP S&P 500 Universe file.

Panel A: Full Sample (1989-2016)

	SPY	VFIAX	IVV
SPY	1	0.030	0.008
VFIAX	0.030	1	0.091
IVV	0.008	0.091	1

Panel B: Early Sample (1989-2000)

	SPY	VFIAX	IVV
SPY	1	−0.118	0.096
VFIAX	−0.118	1	0.170
IVV	0.096	0.170	1

Panel C: Mid Sample (2001-2010)

	SPY	VFIAX	IVV
SPY	1	0.113	0.191
VFIAX	0.113	1	0.361
IVV	0.191	0.361	1

Panel D: Late Sample (2011-2016)

	SPY	VFIAX	IVV
SPY	1	0.005	0.181
VFIAX	0.005	1	0.367
IVV	0.181	0.367	1

Panel E: Crisis Sample (December 2007 to June 2009)

	SPY	VFIAX	IVV
SPY	1	0.125	0.243
VFIAX	0.125	1	0.159
IVV	0.243	0.159	1

Table 3: Statistics on the Growth of Index Funds

All data in this table is sourced from the 2016 ICI Factbook. Index Fund Share of Equity Mutual Fund Assets divides total equity index mutual fund AUM by total equity mutual fund AUM. ETF data includes both 1940-act ETFs and non 1940-act ETFs across all asset classes. Fund-of-funds which invest primarily in other index mutual funds are excluded when counting index mutual funds and from calculations of index mutual fund AUM. Average fees include all US equity mutual funds, including index mutual funds and is not weighted by fund assets under management.

Year	Index Fund Share (%) of Equity Mutual Fund Assets	Number of ETFs	ETF Assets ($B)	Number of Index Mutual Funds	Index Mutual Fund Assets ($B)	US Equity Mutual Fund Average Fees (bps)
2000	9.1	80	66	271	384	99
2001	9.9	102	83	286	371	99
2002	10.7	113	102	313	327	100
2003	11.1	119	150	321	455	100
2004	11.4	152	228	328	554	95
2005	11.2	204	301	322	619	91
2006	11.4	359	423	343	747	88
2007	11.7	629	608	354	855	86
2008	13.6	728	531	360	619	83
2009	13.9	797	777	357	835	87
2010	14.7	923	992	365	1,017	83
2011	16.4	1,134	1,048	382	1,094	79
2012	17.4	1,194	1,337	372	1,311	77
2013	18.4	1,294	1,695	371	1,734	74
2014	20.2	1,411	1,974	383	2,054	70
2015	22.0	1,594	2,100	406	2,207	68

Table 4a: Contemporaneous Relationship between Flows and Returns

Regressions (1) and (2) regress contemporaneous S&P 500 return on adjusted flows and one period change in flows. These regressions take the form

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 AF_t = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6})$$

and

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 F_t$$

respectively. Flow data is constructed from the CRSP mutual funds and monthly returns file. S&P 500 return is taken from the S&P Universe file. All flows (F_t) are scaled by market capitalization. Regressions (3) and (4) regress flows and adjusted flows (as defined above) on one period prior returns.

	Dependent variable:			
	S&P 500 Return (r_t)		Scaled Flows (F_t)	Adjusted Flows
	(1)	(2)	(3)	(4)
Adjusted Flows	0.141^{***}			
	(0.053)			
Scaled Flows (F_t)		0.072		
		(0.050)		
Lagged Return (r_{t-1})			−0.059	0.0003
			(0.061)	(0.057)
Constant	0.007^{***}	0.007^{***}	0.010^{***}	0.0002
	(0.002)	(0.002)	(0.003)	(0.002)
Observations	328	332	332	328
R^2	0.021	0.029	0.003	0.00000
Adjusted R^2	0.018	0.026	−0.0002	−0.003
Residual Std. Error	0.041 (df = 326)	0.041 (df = 330)	0.046 (df = 330)	0.043 (df = 326)
F Statistic	6.952^{***} (df = 1; 326)	10.000^{***} (df = 1; 330)	0.937 (df = 1; 330)	0.00003 (df = 1; 326)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4b: Contemporaneous Relationship between Flows and Returns (w/ Controls)

These regressions test the relationship between S&P 500 current period return and various flow variables. Scaled Flows (F_t) indicate flows into S&P 500 index tracking funds scaled by current S&P 500 market capitalization. Scaled active flows are flows into active managed funds, scaled by current S&P 500 market capitalization. Adjusted flows adjust for the six-month moving average of flows. Regressions take the forms:

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 F_t$$

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 AF_t = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6})$$

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 F_t + \hat{\beta}_2 ActiveFlow_t$$

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6}) + \hat{\beta}_2 (ActiveFlow_t - \sum_{j=t-6}^{t} \frac{ActiveFlow_j}{6})$$

Scaled index flows data is constructed from the CRSP mutual funds and monthly returns file. S&P 500 return is taken from the S&P Universe file. Active flow data comes from the ICI

	Dependent variable:			
	S&P 500 Return (r_t)			
	(1)	(2)	(3)	(4)
Scaled Active Flows			0.096***	
			(0.014)	
Scaled Flows (F_t)	0.072		0.156***	
	(0.050)		(0.048)	
Adjusted Active Flows				0.243***
				(0.024)
Adjusted Flows		0.141***		0.360***
		(0.053)		(0.051)
Constant	0.006***	0.007***	0.001	0.007***
	(0.002)	(0.002)	(0.002)	(0.002)
Observations	333	328	333	328
R^2	0.006	0.021	0.128	0.261
Adjusted R^2	0.003	0.018	0.123	0.257
Residual Std. Error	0.042 (df = 331)	0.041 (df = 326)	0.039 (df = 330)	0.036 (df = 325)
F Statistic	2.073 (df = 1; 331)	6.952*** (df = 1; 326)	24.312*** (df = 2; 330)	57.527*** (df = 2; 325)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 4c: Relationship between Flows and Past Returns

These regressions test the relationship between lagged returns and contemporaneous flows to test for return-chasing behavior over three periods of lag. Scaled flows (F_t) are flows into S&P 500 tracking funds adjusted for market capitalization; adjusted flows are adjusted for the six month trailing moving average of F_t. These regressions take the form

$$\hat{F}_t = \hat{\beta}_0 + \hat{\beta}_1 r_{t-1} + ... + \hat{\beta}_{i-1} r_{t-i-1} + \hat{\beta}_i r_{t-i} \quad i \in (1, 2, 3)$$

$$\hat{AF}_t = \hat{\beta}_0 + \hat{\beta}_1 r_{t-1} + ... + \hat{\beta}_{i-1} r_{t-i-1} + \hat{\beta}_i r_{t-i} \quad i \in (1, 2, 3)$$

Scaled index flows data is constructed from the CRSP mutual funds and monthly returns file. S&P 500 return is taken from the S&P Universe file.

	Dependent variable:			
	Scaled Flows (F_t)		Adjusted Flows	
	(1)	(2)	(3)	(4)
1-Period Lag (r_{t-1})	-0.059	-0.056	0.0003	0.002
	(0.061)	(0.061)	(0.057)	(0.057)
2-Period Lag (r_{t-2})		-0.069		-0.005
		(0.061)		(0.057)
3-Period Lag (r_{t-3})		0.008		0.068
		(0.061)		(0.057)
Constant	0.010^{***}	0.011^{***}	0.0002	-0.0003
	(0.003)	(0.003)	(0.002)	(0.002)
Observations	332	330	328	328
R^2	0.003	0.007	0.00000	0.004
Adjusted R^2	-0.0002	-0.002	-0.003	-0.005
Residual Std. Error	0.046 (df = 330)	0.046 (df = 326)	0.043 (df = 326)	0.043 (df = 324)
F Statistic	0.937 (df = 1; 330)	0.737 (df = 3; 326)	0.00003 (df = 1; 326)	0.472 (df = 3; 324)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 5: Longer-Term Negative Relationship between Pooled Flows and Returns ($n = 6$)

These regressions evaluate the relationship between six months of summed lagged flows into S&P 500 tracking passive funds and other active funds and contemporaneous S&P 500 return. All flows are scaled by total S&P 500 market capitalization. Regressions (1) and (2) weight flows equally over each period. Regressions (3) and (4) geometrically discount older flows. Regressions (1) and (2) take the form

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=t-n}^{t} F_i$$

Regressions (3) and (4) take the form

$$\hat{R}_t = \hat{\beta}_0 + \hat{\beta}_1 \sum_{i=t-n}^{t} \sigma^i \times F_i$$

Scaled index flows data is constructed from the CRSP mutual funds and monthly returns file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	S&P 500 Return (r_t)			
	(1)	(2)	(3)	(4)
Equal-Weighted Pooled Scaled Flows	-0.056^{***}	-0.051^{**}		
	(0.021)	(0.021)		
Equal-Weighted Pooled Control		0.004		
		(0.003)		
Geometric-Weighted Pooled Flows ($\sigma = .8$)			-0.071^{**}	-0.062^{**}
			(0.031)	(0.031)
Geometric-Weighted Pooled Control ($\sigma = .8$)				0.009^{**}
				(0.004)
Constant	0.011^{***}	0.009^{***}	0.010^{***}	0.007^{**}
	(0.003)	(0.003)	(0.003)	(0.003)
Observations	327	327	328	328
R^2	0.022	0.028	0.016	0.031
Adjusted R^2	0.019	0.022	0.013	0.025
Residual Std. Error	0.041 (df = 325)	0.041 (df = 324)	0.041 (df = 326)	0.041 (df = 325)
F Statistic	7.209^{***} (df = 1; 325)	4.743^{***} (df = 2; 324)	5.359^{**} (df = 1; 326)	5.138^{***} (df = 2; 325)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 6: Short Term Relationship between Flows and S&P 500/S&P 400 Return Spread

These regressions evaluate the relationship between flows into S&P 500 tracking funds and active funds and the difference between S&P 500 and S&P 400 return. The S&P 400 consists of the next 400 companies with market capitalizations below the cutoff for the S&P 500. These regressions take the following general forms

$$R_{SP500,t} - R_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1 F_t + \hat{\beta}_2 (R_{SP500,t-1} - R_{SP500,t-1})$$

$$R_{SP500,t} - R_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6}) + \hat{\beta}_2 (R_{SP500,t-1} - R_{SP500,t-1})$$

Scaled index flows data is constructed from the CRSP mutual funds and monthly returns file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	S&P 500 Return - S&P 400 Return			
	(1)	(2)	(3)	(4)
Scaled Flows (F_t)	0.027		0.012	
	(0.025)		(0.026)	
Adjusted Flows		0.027		−0.037
		(0.027)		(0.029)
Scaled Active Flows			−0.019**	
			(0.007)	
Adjusted Active Flows				−0.072***
				(0.013)
Previous Difference in Return	−0.041	−0.037	−0.050	−0.070
	(0.055)	(0.056)	(0.055)	(0.054)
Constant	−0.003***	−0.003**	−0.002*	−0.003***
	(0.001)	(0.001)	(0.001)	(0.001)
Observations	332	328	332	328
R^2	0.005	0.004	0.023	0.087
Adjusted R^2	−0.001	−0.002	0.014	0.079
Residual Std. Error	0.021 (df = 329)	0.021 (df = 325)	0.021 (df = 328)	0.020 (df = 324)
F Statistic	0.805 (df = 2; 329)	0.665 (df = 2; 325)	2.585* (df = 3; 328)	10.309*** (df = 3; 324)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 7: Relationship Between Flows and the S&P 480/S&P 520 Return Spread

These regressions evaluate the relationship between flows into S&P 500 tracking funds and active funds and the difference between the return of the bottom 20 stocks of the S&P 500 and the top 20 stocks of the S&P 400. I construct and reweight the portfolio on a monthly basis using returns data from the CRSP Monthly Stock File and inclusion and deletion data from Compustat. Scaled Flows (F_t) indicate flows into S&P 500 index tracking funds scaled by current S&P 500 market capitalization. Scaled active flows are flows into active managed funds, scaled by current S&P 500 market capitalization. Adjusted flows adjust for the six-month moving average of flows. These regressions take the general forms:

$$R_{SP480,t} - R_{SP520,t} = \hat{\beta}_0 + \hat{\beta}_1 F_t$$

$$R_{SP520,t} - R_{SP480,t} = \hat{\beta}_0 + \hat{\beta}_1 (F_t - \sum_{i=t-6}^{t} \frac{F_i}{6})$$

Scaled index flows data is constructed from the CRSP Mutual Fund and Monthly Stock file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	Return			
	(1)	(2)	(3)	(4)
Scaled Flows	0.029	0.036		
	(0.067)	(0.069)		
Scaled Active Flows		0.009		
		(0.021)		
Adjusted Flows			0.011	0.101
			(0.072)	(0.079)
Adjusted Active Flows				0.100***
				(0.037)
Constant	0.001	−0.0001	0.001	0.001
	(0.003)	(0.004)	(0.003)	(0.003)
Observations	292	292	292	292
R^2	0.001	0.001	0.0001	0.025
Adjusted R^2	−0.003	−0.006	−0.003	0.018
Residual Std. Error	0.055 (df = 290)	0.055 (df = 289)	0.055 (df = 290)	0.055 (df = 289)
F Statistic	0.185 (df = 1; 290)	0.186 (df = 2; 289)	0.023 (df = 1; 290)	3.700** (df = 2; 289)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 8: Relationship Between Flows and S&P 400/S&P 500 Price/Earnings Ratio (n=6)

These regressions summarize the relationship between six months of summed lagged flows into S&P 500 tracking passive funds and other active funds and the contemporaneous difference between the S&P 400 and S&P 500 price-to-earnings ratio. Due to autocorrelation in the dependent variable, I additionally include a one period lagged variable in the regressions. I scale all flows by total S&P 500 market capitalization. Regressions (3) and (4) adjust flows for their six-month moving average. These regressions take the form:

$$PE_{SP500,t} - PE_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP500,t-1} - PE_{SP400,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} F_i$$

$$PE_{SP500,t} - PE_{SP400,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP500,t-1} - PE_{SP400,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} (F_i - \sum_{j=i-6}^{j=i} \frac{F_j}{6})$$

	Dependent variable:			
	S&P 500 P/E-S&P 400 P/E			
	(1)	(2)	(3)	(4)
S&P 500 P/E- S&P 400 P/E (t-1)	0.626***	0.595***	0.636***	0.633***
	(0.088)	(0.089)	(0.088)	(0.088)
Equal-Weighted Pooled Scaled Flows	−5.798	−3.309		
	(4.298)	(4.550)		
Equal-Weighted Pooled Active Flows		6.544		
		(4.191)		
Equal-Weighted Pooled Adjusted Flows			−5.077	−7.668
			(5.008)	(6.059)
Equal-Weighted Pooled Adjusted Active Flows				−3.596
				(4.709)
Constant	−0.767	−1.119	−1.056	−1.351*
	(0.721)	(0.749)	(0.672)	(0.776)
Observations	89	89	88	88
R^2	0.410	0.426	0.404	0.408
Adjusted R^2	0.396	0.406	0.390	0.387
Residual Std. Error	5.408 (df = 86)	5.363 (df = 85)	5.459 (df = 85)	5.473 (df = 84)
F Statistic	29.830*** (df = 2; 86)	21.032*** (df = 3; 85)	28.789*** (df = 2; 85)	19.293*** (df = 3; 84)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 9: Relationship Between Flows and S&P 480/S&P 520 Price/Earnings Ratio (n=6)

These regressions summarize the relationship between six months of summed lagged flows into S&P 500 tracking passive funds and other active funds and the contemporaneous difference between the price-to-earnings ratio of the bottom 20 stocks of the S&P 500 and the top 20 stocks of the S&P 400. Due to autocorrelation in the dependent variable, a one period lagged variable is additionally included in the regressions. All flows are scaled by total S&P 500 market capitalization. Regressions (3) and (4) adjust flows for their six-month moving average. These regressions take the form:

$$PE_{SP480,t} - PE_{SP520,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP480,t-1} - PE_{SP520,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} F_i$$

$$PE_{SP480,t} - PE_{SP520,t} = \hat{\beta}_0 + \hat{\beta}_1(PE_{SP480,t-1} - PE_{SP520,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} (F_i - \sum_{j=i-6}^{j=i} \frac{F_j}{6})$$

	Dependent variable:			
	S&P 480 P/E-S&P 520 P/E			
	(1)	(2)	(3)	(4)
S&P 480 P/E-S&P 520 P/E (t-1)	0.878***	0.876***	0.881***	0.882***
	(0.031)	(0.031)	(0.030)	(0.030)
Equal-Weighted Pooled Scaled Flows	−1.516	−1.410		
	(1.221)	(1.235)		
Equal-Weighted Pooled Active Flows		0.100		
		(0.164)		
Equal-Weighted Pooled Adjusted Flows			−3.273**	−2.591
			(1.461)	(1.724)
Equal-Weighted Pooled Adjusted Active Flows				0.414
				(0.554)
Constant	1.429***	1.397***	1.293***	1.292***
	(0.375)	(0.379)	(0.344)	(0.345)
Observations	272	272	272	272
R^2	0.766	0.766	0.769	0.769
Adjusted R^2	0.764	0.763	0.767	0.767
Residual Std. Error	2.230 (df = 269)	2.233 (df = 268)	2.216 (df = 269)	2.218 (df = 268)
F Statistic	439.598*** (df = 2; 269)	292.503*** (df = 3; 268)	446.971*** (df = 2; 269)	297.680*** (df = 3; 268)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 10: Relationship Between Flows and S&P 480/S&P 520 Book/Market Ratio (n=6)

These regressions summarize the relationship between six months of summed lagged flows into S&P 500 tracking passive funds and other active funds and the contemporaneous difference between the book-to-market ratio of the bottom 20 stocks of the S&P 500 and the top 20 stocks of the S&P 400. Due to autocorrelation in the dependent variable, a one period lagged variable is additionally included in the regressions. All flows are scaled by total S&P 500 market capitalization. Regressions (3) and (4) adjust flows for their six-month moving average. These regressions take the form:

$$BM_{SP480,t} - BM_{SP520,t} = \hat{\beta}_0 + \hat{\beta}_1(BM_{SP480,t-1} - BM_{SP520,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} F_i$$

$$BM_{SP480,t} - BM_{SP520,t} = \hat{\beta}_0 + \hat{\beta}_1(BM_{SP480,t-1} - BM_{SP520,t-1}) + \hat{\beta}_2 \sum_{i=t-n}^{t} (F_i - \sum_{j=i-6}^{j=i} \frac{F_j}{6})$$

	Dependent variable:			
	S&P 480 B/M - S&P 520 B/M			
	(1)	(2)	(3)	(4)
S&P 480 B/M - S&P 520 B/M (t-1)	0.001	−0.004	0.008	0.054
	(0.450)	(0.463)	(0.451)	(0.451)
Equal-Weighted Pooled Scaled Flows	−3.029	−3.081		
	(5.637)	(5.739)		
Equal-Weighted Pooled Active Flows		−0.039		
		(0.784)		
Equal-Weighted Pooled Adjusted Flows			−0.566	−6.883
			(6.929)	(8.187)
Equal-Weighted Pooled Adjusted Active Flows				−3.831
				(2.657)
Constant	0.576	0.602	0.371	0.266
	(1.138)	(1.252)	(1.075)	(1.076)
Observations	283	283	283	283
R^2	0.001	0.001	0.00002	0.007
Adjusted R^2	−0.006	−0.010	−0.007	−0.003
Residual Std. Error	10.653 (df = 280)	10.672 (df = 279)	10.658 (df = 280)	10.638 (df = 279)
F Statistic	0.144 (df = 2; 280)	0.097 (df = 3; 279)	0.003 (df = 2; 280)	0.695 (df = 3; 279)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 11: S&P 500 Historical Inclusions and Deletions by Year

This table summarizes the number of valid S&P 500 inclusion and deletion events in the sample. Valid events have 36 months of return history prior to and following the event and exclude any firms added to or removed from the index due to M&A activity, corporate spin offs, privatization, and bankruptcy. I source data on index constituents from the Compustat North America Database and use the CRSP/Compustat Merged Database to link constituents to return data.

Year	S&P 500 Inclusions	S&P 500 Deletions
1989	3	0
1990	10	3
1991	7	2
1992	4	3
1993	8	3
1994	12	8
1995	18	6
1996	11	11
1997	22	4
1998	26	5
1999	32	1
2000	37	12
2001	20	8
2002	15	10
2003	8	1
2004	11	6
2005	13	2
2006	21	6
2007	27	7
2008	27	10
2009	22	9
2010	12	3
2011	12	8
2012	8	6
2013	8	11
Total	**394**	**145**

Table 12: Relationship Between Flows and Changes in β: Inclusions

These regressions summarize the relationship between flows into S&P 500 tracking funds and other active funds over the 72-month index inclusion period and $\Delta\beta_{SP500}$ and $\Delta\beta_{exSP500}$. The βs are calculated as coefficients of the regression

$$R_{i,t} = \hat{\beta_0} + \hat{\beta_1} R_{SP500,t} + \hat{\beta_2} R_{exSP500,t}$$

$\Delta\beta$s are calculated as the difference in the average βs from the 36 month period before the inclusion event and the 36 month period after the inclusion event. Additionally, Regressions (2) and (4) include the standard deviation of flows over the period. All flows are scaled by S&P 500 market capitalization. These regressions take the form

$$\Delta\beta_{SP500} = \hat{\beta_1} \sum_{i=t-n}^{t+n} F_i + \hat{\beta_2} \sum_{i=t-n}^{t+n} ActiveFlow_i$$

$$\Delta\beta_{SP500} = \hat{\beta_1} \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2}$$

Scaled index flows data is constructed from the CRSP Mutual Fund and Monthly Stock file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	$\Delta\beta_{SP500}$		$\Delta\beta_{exSP500}$	
	(1)	(2)	(3)	(4)
$\sum_t F_t$	39.004		13.242	
	(54.790)		(46.370)	
$\sum_t ActiveFlow_t$	-0.514		-1.536	
	(1.817)		(1.538)	
Standard Dev. of F_t		47.705		95.574
		(334.655)		(287.021)
Constant	0.114	0.355^{**}	-0.278	-0.289^{**}
	(0.149)	(0.138)	(0.127)	(0.119)
Observations	166	166	166	166
R^2	0.014	0.00004	0.035	0.0003
Adjusted R^2	0.004	-0.009	-0.004	-0.006
Residual Std. Error	1.256 (df = 163)	1.261 (df = 164)	1.065 (df = 163)	1.081 (df = 164)
F Statistic	1.146 (df = 2; 163)	0.007 (df = 1; 164)	2.991^{*} (df = 2; 163)	0.047 (df = 1; 164)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 13: Relationship Between Flows and Changes in β: Deletions

These regressions summarize the effect of flows into S&P 500 tracking funds and other active funds over the 72-month index deletion period on $\Delta\beta_{SP500}$ and $\Delta\beta_{exSP500}$. The βs are calculated as coefficients of the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 R_{SP500,t} + \hat{\beta}_2 R_{exSP500,t}$$

$\Delta\beta$s are calculated as the difference in the average βs from the 36 month period before the deletion event and the 36 month period after the deletion event. Additionally, Regressions (2) and (4) include the standard deviation of flows over the period. All flows are scaled by S&P 500 market capitalization. These regressions take the form

$$\Delta\beta_{SP500} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \hat{\beta}_2 \sum_{i=t-n}^{t+n} ActiveFlow_i$$

$$\Delta\beta_{SP500} = \hat{\beta}_1 \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2}$$

Scaled index flows data is constructed from the CRSP Mutual Fund and Monthly Stock file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	$\Delta\beta_{SP500}$		$\Delta\beta_{exSP500}$	
	(1)	(2)	(3)	(4)
$\sum_t F_t$	−215.947		207.986	
	(202.383)		(209.201)	
$\sum_t ActiveFlow_t$	17.630		−20.950	
	(13.479)		(13.933)	
Standard Dev. of F_t		−1,438.635**		1,596.564***
		(557.396)		(574.929)
Constant	−0.220	−0.127	0.362	0.181
	(0.537)	(0.269)	(0.556)	(0.277)
Observations	77	77	77	77
R^2	0.083	0.089	0.092	0.102
Adjusted R^2	0.056	0.076	0.065	0.089
Residual Std. Error	1.732 (df = 71)	1.714 (df = 72)	1.791 (df = 71)	1.768 (df = 72)
F Statistic	3.032* (df = 2; 74)	6.662** (df = 1; 75)	3.386** (df = 2; 74)	7.712*** (df = 1; 75)

Note: *p<0.1; **p<0.05; ***p<0.01

Table 14: Relationship Between Flows and Changes in Illiquidity: Inclusions

These regressions summarize the effect of flows into S&P 500 tracking funds and other active funds over the 72-month index inclusion event period on $\%\Delta ILL_i$. ILL_i is a modified version of Amihud's (2002) measure of illiquidity. $\%\Delta ILL_i$ is calculated as the difference in the average ILL_i in the 36 month period prior to the inclusion event and the 36 month period after the event, divided by the pre-event level of ILL_i in order to control for aggregate liquidity increases over time. Additionally, Regression (2) includes the standard deviation of flows over the period. All flows are scaled by S&P 500 market capitalization. Passive share in Regression (3) is calculated by dividing the total assets in passive S&P 500 funds by S&P 500 market capitalization. Regression (1) uses scaled and active flows over the entire event period and Regression (4) over the 12 months immediately around inclusion. These regressions take the form

$$\%\Delta ILL_i = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \hat{\beta}_2 \sum_{i=t-n}^{t+n} ActiveFlow_i$$

$$\%\Delta ILL_i = \hat{\beta}_1 \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2}$$

$$\%\Delta ILL_i = \hat{\beta}_1 \sum_f \frac{NAV_{f,t}}{IndexCap_t}$$

Scaled index flows data is constructed from the CRSP Mutual Fund and Monthly Stock file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	ΔILL_i			
	(1)	(2)	(3)	(4)
$\sum_t F_t$	28.836**			−17.691
	(13.927)			(21.322)
$\sum_t ActiveFlow_t$	−1.065**			−1.062
	(0.455)			(1.846)
Standard Dev. of F_t		206.184**		
		(98.570)		
Passive Share			−0.007	
			(0.028)	
Constant	−0.642***	−0.557***	−0.454***	−0.438***
	(0.092)	(0.044)	(0.054)	(0.038)
Observations	166	166	181	181
R^2	0.065	0.057	0.0003	0.020
Adjusted R^2	0.047	0.049	−0.005	0.0001
Residual Std. Error	0.374 (df = 107)	0.373 (df = 108)	0.363 (df = 179)	0.396 (df = 167)
F Statistic	3.716** (df = 2; 107)	6.587** (df = 1; 108)	0.061 (df = 1; 179)	1.004 (df = 2; 167)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 15: Relationship Between Flows and Changes in Illiquidity: Deletions

These regressions summarize the effect of flows into S&P 500 tracking funds and other active funds over the 72-month index deletion event period on $\%\Delta ILL_i$. ILL_i is a modified version of Amihud's (2002) measure of illiquidity. $\%\Delta ILL_i$ is calculated as the difference in the average ILL_i in the 36 month period prior to the deletion event and the 36 month period after the event, divided by the pre-event level of ILL_i in order to control for aggregate liquidity increases over time. Additionally, Regression (2) includes the standard deviation of flows over the period. All flows are scaled by S&P 500 market capitalization. Passive share in Regression (3) is calculated by dividing the total assets in passive S&P 500 funds by S&P 500 market capitalization. Regression (1) uses scaled and active flows over the entire event period and Regression (4) over the 12 months immediately around deletion. These regressions take the form

$$\%\Delta ILL_i = \hat{\beta_1} \sum_{i=t-n}^{t+n} F_i + \hat{\beta_2} \sum_{i=t-n}^{t+n} ActiveFlow_i$$

$$\%\Delta ILL_i = \hat{\beta_1} \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2}$$

$$\%\Delta ILL_i = \hat{\beta_1} \sum_f \frac{NAV_{f,t}}{IndexCap_t}$$

Scaled index flows data is constructed from the CRSP Mutual Fund and Monthly Stock file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	ΔILL_i			
	(1)	(2)	(3)	(4)
$\sum_t F_t$	270.673			102.486
	(237.067)			(410.851)
$\sum_t ActiveFlow_t$	−8.529			−27.205
	(7.122)			(31.852)
Standard Dev. of F_t		1,319.072		
		(1,615.641)		
Passive Share			−0.379	
			(0.489)	
Constant	0.893	1.922***	3.370***	2.762***
	(1.604)	(0.746)	(1.005)	(0.765)
Observations	77	77	89	89
R^2	0.020	0.008	0.007	0.009
Adjusted R^2	−0.004	−0.004	−0.005	−0.014
Residual Std. Error	3.384 (df = 74)	3.630 (df = 75)	4.741 (df = 87)	4.428 (df = 86)
F Statistic	0.830 (df = 2; 74)	0.667 (df = 1; 75)	0.600 (df = 1; 87)	0.398 (df = 2; 86)
Note:				*p<0.1; **p<0.05; ***p<0.01

Table 16: Relationship Between Flows and Changes in Illiquidity Loading: Deletions

These regressions examine the effect of flows into S&P 500 tracking funds and other active funds over the 72-month index deletion event period on $\Delta\beta_{ILL}$. ILL_i is a modified version of Amihud's (2002) measure of illiquidity. The βs are calculated as coefficients of the regression

$$R_{i,t} = \hat{\beta}_0 + \hat{\beta}_1 ILL_{i,m}$$

$\Delta\beta_{ILL}$ is calculated as the difference in average β_1 in the 36 month period before the deletion event and the 36 month period immediately following. Additionally, Regression (2) includes the standard deviation of flows over the period. All flows are scaled by S&P 500 market capitalization. Passive share in Regression (3) is calculated by dividing the total assets in passive S&P 500 funds by S&P 500 market capitalization. Regression (1) uses scaled and active flows over the entire event period and Regression (4) over the 12 months immediately around deletion. These regressions take the form

$$\Delta\beta_{ILL} = \hat{\beta}_1 \sum_{i=t-n}^{t+n} F_i + \hat{\beta}_2 \sum_{i=t-n}^{t+n} ActiveFlow_i$$

$$\Delta\beta_{ILL} = \hat{\beta}_1 \sqrt{\frac{1}{2n-1} \sum_{i=t-n}^{t+n} (F_i - \overline{F})^2}$$

$$\Delta\beta_{ILL} = \hat{\beta}_1 \sum_f \frac{NAV_{f,t}}{IndexCap_t}$$

Scaled index flows data is constructed from the CRSP Mutual Fund and Monthly Stock file. S&P 500 return and market capitalization comes from the CRSP S&P Universe file. Active flow data comes from the ICI.

	Dependent variable:			
	$\Delta\beta_{ILL}$			
	(1)	(2)	(3)	(4)
$\sum_t F_t$	13,485.000**			13,485.000**
	(6,342.780)			(6,342.780)
$\sum_t ActiveFlow_t$	530.330			530.330
	(451.904)			(451.904)
Standard Dev. of F_t		110,095.400**		
		(42,712.950)		
Passive Share			12.755*	
			(6.838)	
Constant	−141.264*	−15.459	−6.355	−141.264*
	(76.805)	(19.960)	(14.489)	(76.805)
Observations	77	77	89	89
R^2	0.132	0.084	0.024	0.132
Adjusted R^2	0.108	0.072	0.017	0.108
Residual Std. Error	83.239 (df = 74)	84.913 (df = 75)	84.086 (df = 87)	83.239 (df = 86)
F Statistic	5.419*** (df = 2; 74)	6.644** (df = 1; 75)	3.479* (df = 1; 87)	5.419*** (df = 2; 86)
Note:				*p<0.1; **p<0.05; ***p<0.01

References

Acharya, Viral V, and Lasse Heje Pedersen. "Asset pricing with liquidity risk." *Journal of Financial Economics* 77, no. 2 (2005): 375–410.

Ackman, William. *2015 Letter to Shareholders*. Pershing Square Capital Management, LP, January 2016.

Amihud, Yakov. "Illiquidity and stock returns: cross-section and time-series effects." *Journal of Financial Markets* 5, no. 1 (2002): 31–56.

Amihud, Yakov, Haim Mendelson, Lasse Heje Pedersen, et al. "Liquidity and asset prices." *Foundations and Trends® in Finance* 1, no. 4 (2006): 269–364.

Barberis, Nicholas, Robin Greenwood, Lawrence Jin, and Andrei Shleifer. "X-CAPM: An extrapolative capital asset pricing model." *Journal of Financial Economics* 115, no. 1 (2015): 1–24.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. "Comovement." *Journal of Financial Economics* 75, no. 2 (2005): 283–317.

Brennan, Michael J, and Avanidhar Subrahmanyam. "Market microstructure and asset pricing: On the compensation for illiquidity in stock returns." *Journal of Financial Economics* 41, no. 3 (1996): 441–464.

Buffett, Warren. *2016 Annual Report*. Berkshire Hathaway, Inc, 2017.

Carhart, Mark M. "On persistence in mutual fund performance." *The Journal of Finance* 52, no. 1 (1997): 57–82.

Chen, Honghui, Gregory Noronha, and Vijay Singal. "The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation." *The Journal of Finance* 59, no. 4 (2004): 1901–1930.

Chen, Honghui, Vijay Singal, and Robert F Whitelaw. “Comovement revisited.” *Journal of Financial Economics* 121, no. 3 (2016): 624–644.

Coakley, J, and P Kougoulis. “Comovement and changes to the FTSE 100 Index.” *SSRN Electronic Journal*, 2004.

Ferri, Richard A. *All about index funds: the easy way to get started*. 2nd ed. New York, NY: McGraw-Hill, 2007.

Goetzmann, William N., and Massimo Massa. “Index Funds and Stock Market Growth.” *The Journal of Business* 76, no. 1 (January 2003): 1–28. https://ideas.repec.org/a/ucp/jnlbus/v76y2003i1p1-28.html.

Greenwood, Robin M, and Nathan Sosner. “Trading patterns and excess comovement of stock returns.” *Financial Analysts Journal*, 2007, 69–81.

Grossman, Sanford J, and Joseph E Stiglitz. “On the impossibility of informationally efficient markets.” *The American Economic Review* 70, no. 3 (1980): 393–408.

Gruber, Martin J. “Another puzzle: The growth in actively managed mutual funds.” *The Journal of Finance* 51, no. 3 (1996): 783–810.

Harris, Larry. *Trading and exchanges: Market microstructure for practitioners*. Oxford University Press, USA, 2003.

Hegde, Shantaram P, and John B McDermott. “The liquidity effects of revisions to the S&P 500 index: An empirical analysis.” *Journal of Financial Markets* 6, no. 3 (2003): 413–459.

Hlavac, Marek. *stargazer: Well-Formatted Regression and Summary Statistics Tables.* R package version 5.2. Cambridge, USA: Harvard University, 2015. http://CRAN.R-project.org/package=stargazer.

Investment Company Institute. *2016 ICI Factbook*. 1401 H St., NW, Suite 1200 Washington, DC 20005: Investment Company Institute, May 2016.

Jain, Prem C. “The effect on stock price of inclusion in or exclusion from the S&P 500.” *Financial Analysts Journal*, 1987, 58–65.

Kasch, Maria, and Asani Sarkar. “Is there an S&P 500 index effect?,” 2014.

Kyle, Albert S. “Continuous auctions and insider trading.” *Econometrica: Journal of the Econometric Society*, 1985, 1315–1335.

Lou, Xiaoxia, and Tao Shu. “Price impact or trading volume: Why is the Amihud (2002) illiquidity measure priced?,” 2016.

Lynch, Anthony W., and Richard R. Mendenhall. “New Evidence on Stock Price Effects Associated with Changes in the S&P 500 Index.” *The Journal of Business* 70, no. 3 (1997): 351–383. ISSN: 00219398, 15375374. `http://www.jstor.org/stable/10.1086/209722`.

Madhavan, Ananth. “Market microstructure: A survey.” *Journal of Financial Markets* 3, no. 3 (2000): 205–258.

Malkiel, Burton G. “Returns from investing in equity mutual funds 1971 to 1991.” *The Journal of Finance* 50, no. 2 (1995): 549–572.

McGinty, Tom, Sarah Krouse, and Elliot Bentley. “Index Funds Are Taking Over the S&P 500.” *The Wall Street Journal*, October 2016.

Miller, Ross M, and Miller Risk Advisors. “Measuring the true cost of active management by mutual funds.” *Journal of Investment Management* 5, no. 1 (2007).

Morck, Randall, and Fan Yang. *The Mysterious Growing Value of S&P 500 Membership*. Working Paper, Working Paper Series 8654. National Bureau of Economic Research, December 2011. doi:10.3386/w8654. http://www.nber.org/papers/w8654.

Petajisto, Antti. "Active share and mutual fund performance." *Financial Analysts Journal* 69, no. 4 (2013): 73–93.

Sarr, Abdourahmane, and Tonny Lybek. "Measuring liquidity in financial markets," 2002.

Shleifer, Andrei. "Do demand curves for stocks slope down?" *The Journal of Finance* 41, no. 3 (1986): 579–590.

S&P Dow Jones, ed. *S&P U.S. Indices Methodology*. S&P Global, 2017.

Staer, Arsenio. "Fund flows and underlying returns: The case of ETFs," 2016.

Tufano, Peter, Michael Quinn, and Ryan Taliaferro. "Live prices and stale quantities: T+ 1 accounting and mutual fund mispricing," 2006.

Warther, Vincent A. "Aggregate Mutual Fund Flows and Security Returns." *Journal of Financial Economics* 39, no. 2 (1995): 209–235.

Wermers, Russ. "Mutual fund herding and the impact on stock prices." *The Journal of Finance* 54, no. 2 (1999): 581–622.

Wurgler, Jeffrey. *On the Economic Consequences of Index-Linked Investing*. Working Paper, Working Paper Series 16376. National Bureau of Economic Research, September 2010. doi:10.3386/w16376. http://www.nber.org/papers/w16376.

Wurgler, Jeffrey, and Ekaterina Zhuravskaya. "Does arbitrage flatten demand curves for stocks?" *The Journal of Business* 75, no. 4 (2002): 583–608.