Passive Ownership and Market Efficiency

MARCO SAMMON*

November 1, 2018

ABSTRACT

I document new stylized facts on a decrease in pre-earnings announcement price informativeness. Between 1990 and 2017, pre-earnings cumulative abnormal trading volume declined 10% and the pre-earnings drift declined 22%. Further, earnings days now account for 17% of total annual volatility, up from 3% in 1990. At the firm-level, increases in passive ownership can explain up to 76% of the decline in pre-earnings volume, 20% of the decline in the pre-earnings drift, and 14% of the increase in volatility on earnings days. These results are robust to using only quasi-exogenous variation in passive ownership that arises from S&P 500 index addition, and Russell 1000/2000 index reconstitution. One explanation for decreased efficiency is that passive managers lack strong incentives to gather and consume firm-specific information. Consistent with this mechanism, increases in passive ownership are correlated with fewer analysts covering a stock, decreased analyst accuracy, and fewer downloads of SEC filings.

JEL classification: G12, G14.

*Kellogg School of Management. Email: mcsammon@gmail.com. I would like to thank Scott Baker, Ravi Jagannathan, Robert Korajczyk, Dimitris Papanikolaou and participants at the Kellogg Bag Lunch for helpful comments.

Passive ownership of stocks has grown substantially over the past 30 years. Figure 5 shows that index funds and index ETFs grew from almost zero in 1990, to nearly 40% of total mutual fund and ETF assets in 2017, owning about 10% of the total market capitalization of US firms.[1]

There is an ongoing debate in the academic literature and popular press about the effect of growing passive ownership on market efficiency. A prominent view, backed by hedge fund manager Seth Klarman[2], is that passive ownership has made prices less informative. Based on theory alone, however, this conclusion is not obvious. In fact, Glosten, Nallareddy, and Zou (2016) explain that ETFs could make prices more informative by increasing liquidity for the underlying stocks and expanding the number of shares available for short selling. Moreover, passive ownership is still relatively small, holding only 10% of the total market capitalization of US stocks. At this low level, there may be no observable change in price informativeness. Without clear predictions from theory, empirical work is needed to understand the effect of growing passive ownership on market efficiency.[3]

In this paper, I measure price efficiency as the incorporation of earnings information into prices before the announcement date. In particular, I utilize three measures of pre-earnings announcement price informativeness: (1) Pre-earnings cumulative abnormal volume: the total abnormal volume in the month before earnings announcements (2) Pre-earnings drift: the ratio of the returns on earnings days to the cumulative returns leading up to earnings days (3) Earnings day share of annual volatility: the squared returns on the four quarterly earnings announcement dates divided by the total squared returns in each calendar year.

Over the past 30 years, pre-earnings volume, and the pre-earnings drift have been trending downward, while the share of volatility on earnings days has been trending upward. These trends are consistent with a decrease in the share of earnings information incorporated into prices in advance of the announcement, and thus a decrease in the informational efficiency of the market.

Using data on all US stocks, and all US mutual fund holdings, I test whether these measures of informativeness are related to the fraction of a firm's shares owned by passive funds. Across all three measures, there is a negative relationship between passive ownership and pre-earnings price informativeness. This reduced-form result, however, does not rule out the possibility that unobserved or omitted factors are driving both the increases in passive ownership and decreases in price informativeness.

To address this potential omitted variable bias, I exploit the timing of S&P 500 index additions, as well as Russell 1000/2000 index reconstitutions. Within a narrow band of comparable firms, changes in passive ownership associated with index rebalancing are plausibly uncorrelated with firm

[1]Passive ownership is defined as all index funds, all ETFs that are not actively managed, and all mutual funds with "index" in the name. Index funds are identified using the index fund flag in the CRSP mutual fund data. Total equity mutual fund and ETF assets is the sum of all stock holdings in the Thompson S12 data. Total market capitalization includes all CRSP firms.

[2]See, for example, Klarman's article in the New York Times and his 2016 Year-End Letter.

[3]The growth of index ETFs and index mutual funds does not necessarily imply an increase in passive ownership. Over time, institutional investors like university endowments and insurance companies have substituted their holdings of individual stocks for index funds. If these institutions should be classified as passive investors, the growth of index funds has come partly at the expense of decreased passive ownership of individual stocks.

fundamentals. These quasi-exogenous increases in passive ownership are also negatively correlated with pre-earnings price informativeness.

One potential mechanism through which passive ownership may decrease price informativeness is that passive managers, as well as investors in passive funds, lack strong incentives to gather and consume firm-specific information.[4] Passive funds trade on mechanical rules, such as S&P 500 index membership (SPY), or the 100 lowest volatility stocks in the S&P 500 (SPLV). These rules are based on public information, and thus do not require accurate private forecasts of firm fundamentals. I find a negative relationship between increases in passive ownership and the number of analysts covering a stock, the accuracy of analyst forecasts, and downloads of SEC filings both across stocks, and across time for the same stocks.

My results contribute to the literature in several dimensions. First, my empirical results are consistent with predictions from models of asymmetric information. In Grossman and Stiglitz (1980), prices become less informative as the share of uninformed investors increases. I show that increases in passive ownership lead to a lower pre-earnings drift, suggesting less earnings information is being incorporated into prices before the announcement date. Wang (1994) predicts that less informative prices before information releases would lead to lower trading volume through a fear of adverse selection. I find that higher passive ownership leads to less pre-earnings trading. In Buffa, Vayanos, and Woolley (2014), prices of stocks with high demand by buy-and-hold investors, whose behavior is similar to passive managers in practice, respond more to expected cashflow shocks. I find that stock prices of firms with high passive ownership respond more to a given level of earnings surprise than firms with low passive ownership.

I also contribute to the empirical literature on the effects of growing ETFs and passive ownership. Israeli, Lee, and Sridharan (2017) show that ETFs decrease the information content of the underlying securities' prices. I sharpen this result, providing causal evidence on the effect of increased passive ownership through S&P 500 index addition and Russell 1000/2000 reconstitution. My paper also goes beyond just ETFs, as my definition of passive ownership includes index funds. In 2000, ETFs were less than half of total passive ownership. Although by 2017, ETFs had grown to 70% of passive ownership, index mutual funds are still large, owning about 3% of the US equity market. From an incentives standpoint, index funds and ETFs should have a similar effect on information-gathering, as they both encourage trading baskets of securities rather than individual stocks.

Glosten et al. (2016) find that ETFs increase the incorporation of systematic news into prices for otherwise information deficient stocks. While ETFs may increase the rewards for gathering sector or market-level information, I find they decrease the incentives for gathering firm-specific information.

The paper is organized as follows: Section I shows that average pre-earnings-announcement price informativeness has decreased over the last 30 years. Section II presents a model where

[4] As shown in Glosten et al. (2016), ETFs may increase the benefits of gathering index-level or sector-level information for any given firm. My claim is a narrow statement about firm-level *idiosyncratic* information.

increases in passive ownership can decrease price informativeness. Section III shows the reduced-form relationship between high passive ownership and decreased information in prices. Section IV provides evidence that investors are gathering less firm-specific information for stocks with high passive ownership. Section V uses index rebalancing to create a causal link between passive ownership and decreased price informativeness. Section VI concludes.

I. New Stylized Facts

In this section, I show that three measures of pre-earnings-announcement price informativeness have declined over the past 30 years.

A. Decline of Pre-Earnings Volume

Fact 1: Volume before earnings announcements has declined.

Let t denote an earnings announcement date, or the next trading date if earnings are announced when markets are closed. Define pre-earnings cumulative abnormal volume *per trading day*, for firm i, from time $t-22$ to $t-j$ as:

$$\overline{CAV_{i,j,t}} = \frac{\sum_{\tau=-22}^{-j} AV_{i,t+\tau}}{23-j} \tag{1}$$

Where abnormal volume, $AV_{i,\tau}$, is volume relative to historical average volume over the past year:

$$AV_{i,\tau} = \frac{V_{i,t+\tau}}{\overline{V_{i,t-22}}} = \frac{V_{i,t+\tau}}{\sum_{\tau=252}^{1} V_{i,t-\tau-21}/252} \tag{2}$$

In Equation 2, $V_{i,t+\tau}$ is total daily volume in CRSP. Historical average volume, $\overline{V_{i,t-22}}$, is fixed at the beginning of the 22-day window before earnings are announced to avoid mechanically amplifying drops in volume.

I run the following regression with daily data to measure abnormal volume around earnings announcements:

$$\overline{CAV_{i,j,t}} = \alpha + \sum_{\tau=-21}^{0} \beta_\tau \mathbf{1}_{\{j==-\tau\}} + \text{Fixed Effects} + e_{i,j,t} \tag{3}$$

The main right hand side variables are a set of 22 indicators for days relative to the earnings announcement. For example, $\mathbf{1}_{\{j==15\}}$ is equal to one 15 trading days before the next earnings announcement, and zero otherwise. The regression also includes firm, year and day-of-the-week fixed effects.[5] The regression includes all firms that can be matched between CRSP and I/B/E/S. I run this regression for 4 time periods: (1) 1994-1999 (2) 2000-2004 (3) 2005-2010 (4) 2011 to 2017.

[5]Year fixed-effects are included to account for level differences in pre-earnings volume across years within each period. All results are robust to removing the year fixed effects.

Figure 1 plots the estimates of $\beta_{\{j==-\tau\}}$.

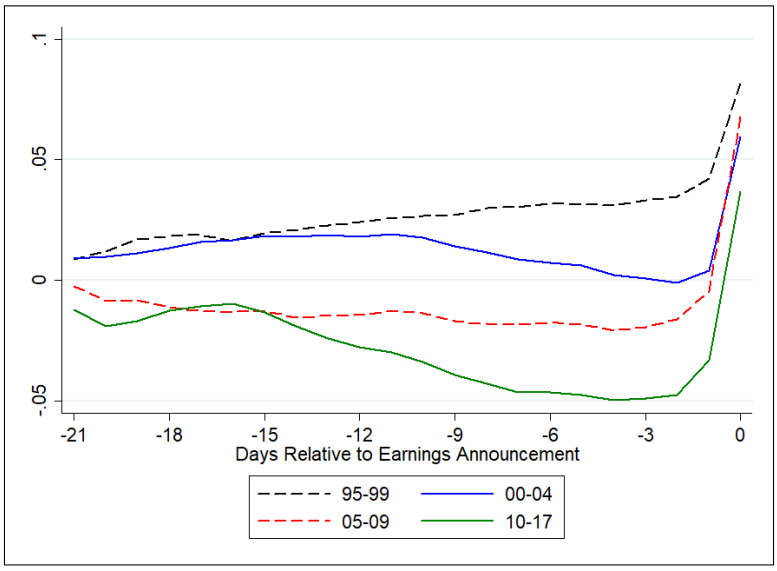

Figure 1. Decline of Pre-Earnings Volume. Plot of $\beta_{\{j==-\tau\}}$ estimated from the regression:

$$\overline{CAV_{i,j,t}} = \alpha + \sum_{\tau=-21}^{0} \beta_\tau \mathbf{1}_{\{j==-\tau\}} + \text{Fixed Effects} + e_{i,j,t}$$

The units of the left-hand-side variable are average cumulative abnormal volume per trading day. The regression includes year and day-of-the-week fixed effects. For the 1995-1999 period, $\beta_{\{j==1\}} = .042$, while for the 2010-2017 period, $\beta_{\{j==1\}} = -0.033$. The total difference of 0.075 is a decline of $0.075 \times 22 \approx 1.65$ days worth of pre-earnings abnormal volume.

The regression coefficients imply that between the late 90's and the present day, there was a decline of about 1.65 days worth of abnormal volume over the 22-day window before earnings announcements.[6] In the raw data, average $\overline{CAV_{i,j,t}}$ was 1.15 between 1995 and 1999, and 1.03 between 2010 and 2017, implying a 10.4% decline in pre-earnings trading.[7]

B. Pre-Earnings Drift

Fact 2: The pre-earnings drift has declined.

Let $E_{i,t}$ denote earnings per share for firm i in quarter t in the I/B/E/S Unadjusted Detail File. Define standardized unexpected earnings (SUE) as in Novy-Marx (2015): The year-over-year (YOY) change in earnings, divided by the standard deviation of YOY changes in earnings over the past 8 quarters.

$$SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})} \tag{4}$$

[6] Section D of the appendix motivates the choice of a 22-trading-day window before the announcement.

[7] This implies a decline of -0.14 $\times$ 22 $\approx$ 2.30 days worth of pre-earnings abnormal volume. This does not account for the firm, year and day-of-the-week fixed effects in Regression 3, which may explain why the estimated effect is larger.

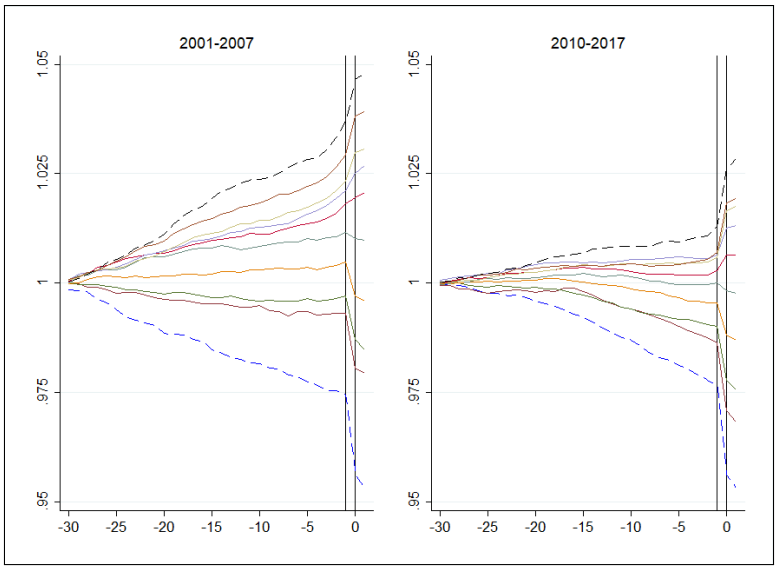

Figure 2. Decline of Pre-Earnings Drift by SUE Decile. Each quarter, I sort firms into deciles on:

$$SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$$

Each line represents the cross-sectional average cumulative market-adjusted return by SUE decile. The black dashed line represents the average for firms with the most positive earnings surprises, while the blue dashed line represents the average for firms with the most negative earnings surprises.

Define market-adjusted returns, $r_{i,t}$, as in Campbell, Lettau, Malkiel, and Xu (2001): the difference between firm i's excess return and the return on the market factor from Ken French's data library. Each quarter, I sort firms into deciles of SUE, and calculate the average cumulative market-adjusted returns over the 30 days prior to the earnings announcement. Figure 2 shows the average pre-earnings cumulative returns by SUE decile for two different time periods: 2001-2007 and 2010-2017. The black dashed line represents the average for firms with the most positive earnings surprises, while the blue dashed line represents the average for firms with the most negative earnings surprises. Between 2010 and 2017, firms in each decile move less before earnings days, relative to the returns on earnings days themselves, than between 2001 and 2007. This decline in pre-earnings drift is even stronger when comparing to the pre-2001 period, but that may be due to Regulation Fair Disclosure (Reg FD), implemented in August, 2000, which limited firms' ability to selectively disclose earnings information before it was publicly announced.

The apparent decline in the pre-earnings drift in Figure 2 could be driven by differences in overall return volatility or average returns between the two time periods. To quantify the decline of the pre-earnings drift, I create a drift magnitude variable designed to capture the share of earnings information incorporated into prices before the announcement date. Let t denote an earnings announcement date. Define the pre-earnings drift for firm i as the cumulative market-adjusted

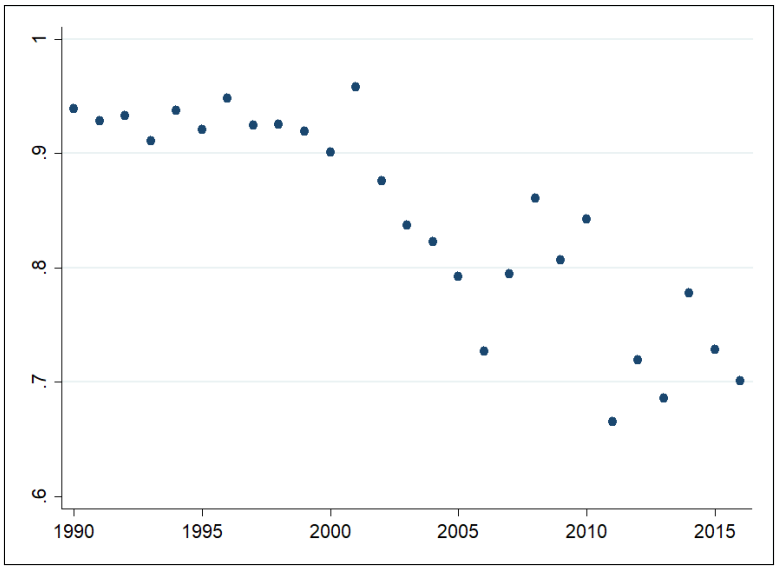

Figure 3. Decline of Average Pre-Earnings Drift. This figure plots the cross-sectional average of the pre-earnings drift, $DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, by year. A value near 1 implies most earnings information is incorporated in prices before the announcement date, while lower values denote less informative pre-earnings prices.

return from $t-30$ to $t-1$, divided by the cumulative returns from $t-30$ to t:

$$DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}} \tag{5}$$

The pre-earnings drift will be near one when the earnings day move is small relative to the cumulative pre-earnings returns. $DM_{i,t}$ will be near zero when the earnings-day return is large, and in the same direction as the cumulative returns before the earnings day. $DM_{i,t}$ will be negative when the earnings day return is a reversal relative to the pre-earnings return. One concern with the definition of pre-earnings drift is that $r_{i,(t-30,t)}$ may take values near zero, leading to huge values of $DM_{i,t}$. Because $r_{i,(t-30,t)}$ is a cumulative return over 30 days, fewer than 1% of all observations are smaller than 10 basis points in absolute value. To further alleviate this concern, I Winsorize $DM_{i,t}$ at the 1% and 99% level by year.[8] Figure 3 shows the cross-sectional average value of $DM_{i,t}$ by year. The pre-earnings drift decreased by about 22% between 1990 and 2017.

C. Earnings Day Volatility

Fact 3: The share of total volatility occurring on earnings days has increased.

Define the quadratic variation share (QVS) for firm i in year t as:

$$QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2 \tag{6}$$

[8]Section E of the Appendix presents alternative definitions of the pre-earnings drift, and further motivates my specification for $DM_{i,t}$.

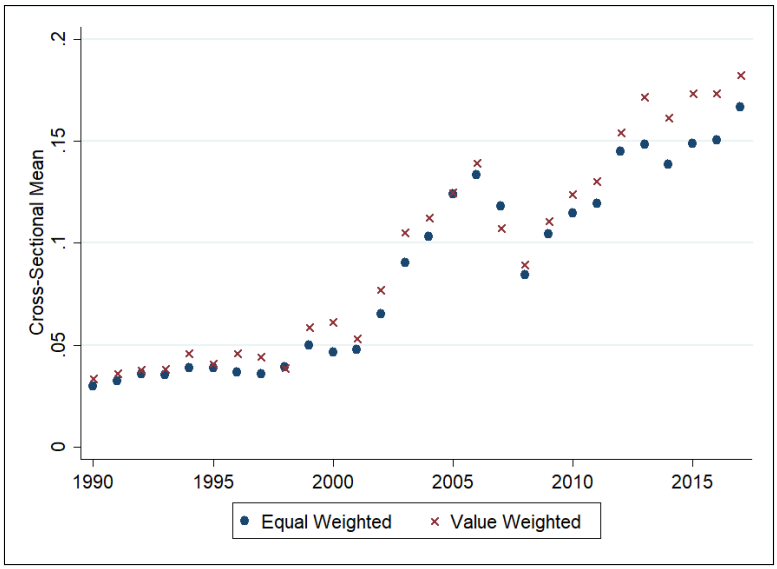

Figure 4. Increase in Earnings Day Volatility. This figure plots the share of market-adjusted quadratic variation occurring on earnings days. For firm i in year t the quadratic variation share (QVS) is defined as: $QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, where r denotes a market-adjusted daily return. The numerator sums over the 4 quarterly earnings days in year t, while the denominator includes all days in calendar year t. The equal-weighted specification is the cross-sectional average for QVS, while the value-weighted specification weights firms in proportion to their market capitalization in year $t-1$.

where r denotes a market-adjusted daily return. The numerator is the sum of squared returns on the 4 quarterly earnings days in year t, while the denominator is the sum of squared returns for all days in year t. Earnings days make up roughly 1.6% of trading days, so values of $QVS_{i,t}$ larger than 0.016 imply that earnings days account for a disproportionately large share of total volatility. Figure 4 shows the cross-sectional average of $QVS_{i,t}$ by year for all CRSP firms that can be matched to 4 non-missing earnings days in a given year in I/B/E/S. Average QVS increased from 3.0% in 1990 to 16.7% in 2017.

D. Discussion

These downward trends in market efficiency could be unrelated to the information released on earnings days. To test this, I reconstruct the time-series averages of the pre-earnings volume, drift and share of volatility on earnings days, except I randomly assign one day each quarter for each firm to be an earnings date. In Section C of the Appendix, Figure 9 shows that there is no drop in volume before the placebo earnings dates. Figure 10 shows that there is no downward trend in the pre-earnings drift for the placebo earnings dates. Figure 11 shows there is no upward trend in the share of volatility on the placebo earnings dates. These results confirm that the changes to price informativeness are specific to earnings days.

As an additional check, Section I of the Appendix examines volume, drift and volatility around Federal Open Market Committee (FOMC) meeting dates instead of placebo earnings dates. I find

no trend toward decreased efficiency in the incorporation FOMC meeting information. This further corroborates that the reduction in efficiency only applies to firm-specific information.

II. Theoretical Predictions

In this section, I examine the effect of increasing passive management in a Grossman and Stiglitz (1980)-style model of asymmetric information. The full details of the model are in Section J of the Appendix.

A. Model Setup

Consider a two period model with a single risky asset that pays a liquidating dividend of d in period 1, with total supply S. Assume d is normally distributed with mean $\overline{d}$ and variance σ_d^2. There are three types of agents: N_I of them are informed, and get a private signal $s = d + \epsilon$ in period 0, where ϵ is normally distributed with mean zero and variance σ_ϵ^2. N_U are uninformed, and only learn about the liquidating dividend through the price. Both the informed and uninformed have exponential utility over period 1 consumption, with coefficients of absolute risk aversion α_I and α_U.

The third type of agents are noise traders who buy and sell u shares with no regard for the price, where u is normally distributed with mean zero and variance σ_u^2. This implies that the effective supply of the asset is $S + u$. The presence of noise traders prevents the uninformed agents from perfectly learning the informed agents' signal from the price in a rational expectations equilibrium.

Prices are less informative if the conditional variance of the dividend given the price, $\sigma_{d|p}^2$, is high, so define price informativeness as $\frac{1}{\sigma_{d|p}^2}$. If $\alpha_I = \alpha_U = \alpha$, then:

$$informativeness = \frac{\sigma_d^2 + \sigma_\epsilon^2 + \left(\frac{\alpha\sigma_\epsilon^2}{N_I}\right)^2 \sigma_u^2}{\sigma_d^2\left(\sigma_\epsilon^2 + \left(\frac{\alpha\sigma_\epsilon^2}{N_I}\right)^2 \sigma_u^2\right)} \tag{7}$$

Prices will be more informative if there are more informed investors, and prices will be less informative if the variance of noisy demand, σ_u^2, is high.

B. Null and Alternative Hypotheses

Period 1 can be viewed as an earnings announcement date, where uncertainty about firm fundamentals (the terminal dividend) is resolved. To predict the effect of rising passive ownership on pre-earnings (period 0) price informativeness, we need to determine how passive managers fit into the model. In some ways, passive managers act like uninformed agents: If there were no noise traders, uninformed agents would become index funds, buying a value-weighted portfolio of the available risky assets. In other ways, passive managers act like noise traders: They allocate based

on mechanical rules like index membership, which does not depend on the fundamental information conveyed by the price.

Null Hypothesis: There will be no relationship between passive ownership and pre-earnings price informativeness

Passive ownership is only 10% of the market, and rarely holds more than 15% of any individual stock. Regardless of whether passive managers act as uninformed investors or noise traders, it is possible they are too small to have any observable effect on price informativeness.

Alternative Hypothesis I: There will be a negative relationship between passive ownership and pre-earnings price informativeness

Mauboussin, Callahan, and Majd (2017) show that the large inflows into passive funds have come at the same time as large outflows from active funds. If we view active managers as informed, and passive managers as uninformed, then the model predicts that prices will become less informative. We could observe a similar effect if we believe passive managers are better modeled as noise traders. If increases in the number of noise traders imply an increase in σ_u^2, average informativeness would go down, holding all else equal.

Alternative Hypothesis II: There will be a positive relationship between passive ownership and pre-earnings price informativeness

The growth in passive ownership has not just been an increase in assets under management, but also an increase in span. The availability of a diverse set of ETFs and index funds makes it easier for active managers to hedge market risk or sector risk, and be more aggressive when betting on their private signals. One way to model this is to allow $\alpha_I \neq \alpha_U$. Decreasing α_I makes prices more informative. Outside the model, ETFs could also increase price informativeness by increasing liquidity or making it easier to short sell shares, as discussed in Glosten et al. (2016).

C. Limitations

Even if we correctly map passive ownership into the model, it is not obvious how to classify the managers who lost assets under management to passive funds. If there has been a decline in informed investors, prices will become less informative, regardless of whether passive managers should be modeled as uninformed investors or noise traders. Alternatively, if the money came from uninformed investors or noise traders, the effect on price informativeness is ambiguous.

Outside changing the mix of agents, there have been major shifts in financial markets between 1990 and 2017 that could also affect market efficiency. Azar, Schmalz, and Tecu (2018) show that ownership has become more concentrated, which may decrease competition among publicly-traded firms. There has also been a dramatic increase in algorithmic trading activity. Weller (2017) shows that ATs can decrease the returns to gathering information, and make markets less efficient. Finally,

there have been changes in regulation, including Sarbanes-Oxley and Regulation Fair Disclosure, which have changed the nature and timing of information releases to investors.

Even ignoring all these concurrent changes to financial markets, the model does not have a clear prediction for the effect of increasing passive ownership, so empirical work is needed to understand passive managers' roles and determine which effect dominates in practice.

III. Reduced-Form Evidence

In this section, I show the reduced-form relationships between increases in passive ownership and declines in pre-earnings volume, declines in pre-earnings drift and increases in the share of volatility on earnings days. I also address competing hypotheses for decreased pre-earnings price informativeness, including the rise of algorithmic trading and Regulation Fair Disclosure.

A. Data and Definitions

Passive ownership is defined as all index funds, all ETFs that are not actively managed, and all mutual funds with "index" in the name. Index funds are identified using the index fund flag in the CRSP mutual fund data. All quarterly fund holdings are from the Thompson S12 data. I use the WRDS MF LINKS database to connect the funds identified as passive in CRSP with the holdings in the Thompson S12 data. Given that holdings are only updated quarterly, I linearly interpolate shares held between S12 filing dates.[9] If a security never appears in the S12 data, I assume the passive ownership share is zero. Figure 5 shows that passive ownership increased from almost zero in 1990, to nearly 40% of total mutual fund and ETF assets in 2017, owning about 10% of the total market capitalization of US firms.

I believe this is a conservative definition of passive ownership, as there are institutional investors which track broad market indices, but are not classified as mutual funds, and do not show up in the S12 data. Further, as discussed in Mauboussin et al. (2017), there has been a rise of closet indexing among self-proclaimed active managers, which will also not show up in my definition of passive management.[10]

All return and daily volume data are from CRSP. I merge CRSP to I/B/E/S (IBES) on CUSIP, or historical CUSIP (ncusip) if available. I use the earnings release times in IBES to identify the first time market participants could trade on earnings information during normal market hours. If earnings are released before 4:00 PM EST between Monday and Friday, that day will be labeled as the effective earnings date. If earnings are released on or after 4:00 PM EST between Monday and Friday, the next trading day will be labeled as the effective earnings date. If earnings are released

[9] All results are robust to fixing shares held to their end of quarter value, rather than interpolating between quarters.

[10] To include closet indexers in the definition of passive management, start by defining passive ownership as above, and then add passive share from Cremers and Petajisto (2009) times the number of shares held by each active mutual fund. Almost all results in the paper are qualitatively similar using this alternative definition of passive management.

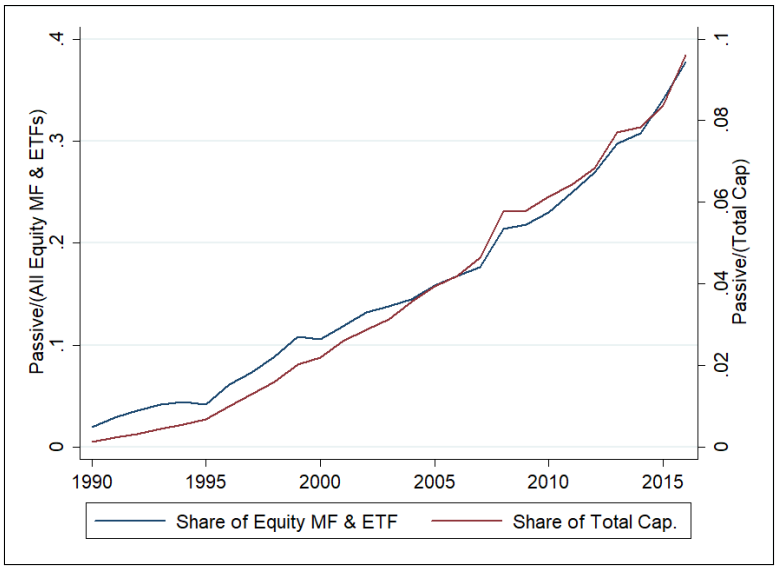

Figure 5. The Rise of Passive Ownership: 1990-2017. Passive ownership is defined as all index funds, all ETFs that are not actively managed, and all mutual funds with "index" in the name. Index funds are identified using the index fund flag in the CRSP mutual fund data. Total equity mutual fund and ETF assets is the sum of all stock holdings in the Thompson S12 data. Total market capitalization includes all CRSP firms.

over the weekend, or on a trading holiday, the next trading date will be labeled as the effective earnings date.

I define quarterly earnings per share as the "value" variable from the IBES unadjusted detail file.[11] All other firm fundamental information is from Compustat.

Total institutional ownership is the sum of shares held by all 13-F filing institutions. Institutional ownership is merged to CRSP on CUSIP, or historical CUSIP if available. If a CUSIP never appears in the 13-F data, institutional ownership is assumed to be zero.

B. Pre-Earnings Volume

Define pre-earnings cumulative abnormal volume *per day* in the 22-day window before earnings is announced as:

$$CAV_{i,t} = \frac{\sum_{\tau=-22}^{-1} AV_{i,t+\tau}}{22} \tag{8}$$

where abnormal volume, AV, is defined in Equation 2. I run the following regression with quarterly data to measure the relationship between declines in pre-earnings volume and increases in passive ownership:

$$\Delta_{(t,t-n)} CAV_{i,t} = \alpha + \beta \Delta_{(t,t-n)} Passive_{i,t} + \gamma X_{i,t-n} + \text{Fixed Effects} + e_{i,t} \tag{9}$$

[11] All results are similar when using Diluted Earnings Per Share Excluding Extraordinary Items (EPSFXQ) in Compustat.

$\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. I only look at year-over-year changes to avoid differences in volume before annual earnings announcements and quarterly announcements or seasonal effects. Controls in $X_{i,t-n}$ include lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership, calculated as the sum of holdings in the 13-F filings. I also condition on the growth in market capitalization from $t-n$ to t.[12] I condition on market capitalization and growth of market capitalization because most of the increase in passive ownership has been in large stocks, and I want to rule out a firm-size effect driving my results. Fixed effects include 2-digit SIC industry, year and security (permno). All standard errors are clustered at the firm/year level.[13]

Given that passive ownership is slow moving, I examine changes in passive ownership and pre-earnings volume over 1, 3 and 5-year horizons. Table I contains the regression results. To interpret the magnitude of the coefficient on $\Delta_{(t,t-n)}Passive_{i,t}$: (1) The 25th percentile of passive ownership share in my sample is 0, while the 75th percentile is around 0.1 (10%) (2) The units of the left hand side variable are abnormal volume per day over the 22 days leading up to the earnings announcement (3) The decline in pre-earnings volume between the late 90's and present day was 1.65 trading days.

The coefficients on $\Delta_{(t,t-n)}Passive_{i,t}$ in the 3-year specification with firm fixed effects implies that moving from the 25th to the 75th percentile of passive ownership would explain (-0.57 $\times$ 22 $\times$0.10 $\approx$ -1.25, from a total decline of -1.65) about 76% of the decline in pre-earnings abnormal volume.[14]

C. Pre-Earnings Drift

Define the pre-earnings drift magnitude as in Equation 5: $DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$. I run the following regression with quarterly data to measure the relationship between the pre-earnings drift and passive ownership.

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=2}^{10} \phi_j \mathbf{1}_{SUEdecile=j} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \tag{10}$$

Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. The regression also includes a set of indicator variables for SUE decile within a given quarter. These are included because as shown in Figure 2, the size of the drift depends on the eventual earnings surprise. Fixed effects include 2-digit SIC industry, year and firm. Standard errors are clustered at the firm/year level.

[12] The results are unchanged if all first differences are replaced with log growth rates.

[13] All results are similar when computing standard errors with panel Newey-West, setting the lags equal to 1.5x the number of overlapping observations, rounding up to the nearest integer.

[14] If we instead consider the raw decline in pre-earnings cumulative abnormal volume of -2.30 trading days, about 55% can be explained by the increase in passive ownership.

		1-year		3-year		5-year	
		(1)	(2)	(1)	(2)	(1)	(2)
Increase in Passive Ownership	1-year	-1.263*	-1.090***				
		(0.651)	(0.215)				
	3-year			-0.936***	-0.573***		
				(0.309)	(0.160)		
	5-year					-1.041**	-0.545***
						(0.421)	(0.174)
Observations		272,609	272,609	191,654	191,654	139,537	139,537
R-Squared		0.009	0.010	0.007	0.008	0.025	0.025
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

Table I Passive Ownership and Pre-Earnings Volume. Estimates of β from:

$$\Delta_{(t,t-n)}CAV_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$ is average pre-earnings cumulative abnormal volume *per day* over the 22 days leading up to the earnings announcement. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Change in passive ownership is expressed as a decimal, so $0.01 = 1\%$ increase. Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry, year and firm. Specification (1) includes all firm-level controls, plus industry and year fixed effects. Specification (2) adds includes firm fixed effects. All standard errors are clustered at the firm/year level.

The regression results are in Table II. Figure 3 shows that DM declined by about 0.2 between 1990 and 2017. The estimated coefficient on $Passive$ in the specification with firm fixed effects implies that moving from the 25th (0) to the 75th percentile (0.1) of passive ownership would explain about ($0.44 \times 0.10 \approx 0.04$, from a total decline of 0.20) about 20% of the decline in pre-earnings drift.[15]

$DM_{i,t}$ is Winsorized at the 1% and 99% level by year to minimize the effect of values of $r_{i,(t-30,t)}$ near zero. Winsorizing $DM_{i,t}$ at the 5% and 95% level increases the statistical significance of the coefficient on $Passive_{i,t}$, as this further reduces the standard errors. Year-over-year changes in drift are volatile, so running the regression in first differences yields similar point estimates, but larger standard errors.

D. Share of Volatility on Earnings Days

Define the earnings day share of annual volatility as in Equation 6: $QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$. I run the following regression with annual data to measure the relationship between changes in

[15] One potential problem with this definition of pre-earnings drift is that the average post-earnings drift has declined (see e.g. McLean and Pontiff (2016)). If the returns that historically were realized after earnings announcements moved to the earnings days themselves, there could be a mechanical decrease in DM, as the average magnitude of $r_{i,t}$ increased. In unreported results, I find that regression estimates are similar, when defining drift as: $\widetilde{DM}_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t+k)}}$ where $k \in \{1,2,3\}$, which confirms this pattern is specific to the pre-earnings drift. Section E of the Appendix shows that the post-earnings drift is larger for firms with higher passive ownership, suggesting that high passive ownership slows the absorption of earnings information after it is released.

	Pre-Earnings Drift	
	(1)	(2)
Passive Ownership	-0.329**	-0.436**
	(0.157)	(0.177)
Observations	413,328	413,328
R-Squared	0.003	0.002
Firm Controls	Yes	Yes
SUE	Yes	Yes
Industry/Year FE	Yes	Yes
Firm FE	No	Yes

Table II Passive Ownership and Pre-Earnings Drift. Table with estimates of β from:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=2}^{10} \phi_j \mathbf{1}_{SUEdecile=j} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, which is the ratio of the cumulative returns in the 30 days leading up to the earnings day, relative cumulative return in the 30 days up to and including the earnings day. Passive ownership is expressed as a decimal, so 0.01 = 1% of shares outstanding held by passive funds. SUE deciles are formed each quarter. $DM_{i,t}$ is Winsorized at the 1% and 99% levels. Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry, year and firm. Specification (1) includes all firm-level controls, plus industry and year fixed effects. Specification (2) adds includes firm fixed effects. All standard errors are clustered at the firm/year level.

earnings day share of annual volatility, and changes in passive ownership:

$$\Delta_{(t,t-n)} QVS_{i,t} = \alpha + \beta \Delta_{(t,t-n)} Passive_{i,t} + \gamma X_{i,t-n} + \text{Fixed Effects} + e_{i,t} \tag{11}$$

Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. I also condition on change in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry, year and firm. Standard errors are clustered at the firm/year level.

The regression results are in Table III. Figure 4 shows that QVS increased by about 0.12 between 1990 and 2017. The coefficients on $\Delta_{(t,t-n)} Passive_{i,t}$ in the 5-year specification without firm fixed effects implies that moving from the 25th (0) to the 75th percentile (0.1) of passive ownership would explain about ($0.17 \times 0.10 \approx 0.02$, from a total increase of 0.14) about 14% of the average increase in share of annual QV on earnings days, depending on the specification considered.

E. Placebo Tests

To confirm that my results are specific to earnings days, I re-run the three reduced-form regressions, except I select dates between the actual earnings days to represent placebo earnings dates. For example, if a firm released earnings on 12/31/2017, I would select the trading day closest to 11/15/2017 as the placebo earnings date. Appendix Tables XII, XIII and XIV compare the original

		1-year		3-year		5-year	
		(1)	(2)	(1)	(2)	(1)	(2)
Increase in Passive Ownership	1-year	0.0983**	0.0551*				
		(0.039)	(0.030)				
	3-year			0.103***	0.0502*		
				(0.034)	(0.026)		
	5-year					0.171***	0.0883***
						(0.042)	(0.027)
Observations		79,253	79,253	60,578	60,578	48,484	48,484
R-Squared		0.015	0.015	0.052	0.047	0.077	0.065
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

Table III Passive Ownership and Earnings Day Share of Volatility. Table with estimates of β from:

$$\Delta_{(t,t-n)}QVS_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + \gamma X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, which is the ratio of the squared returns on the 4 quarterly earnings announcement days, relative to the squared returns on all days in year t. QVS takes values in [0,1]. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry, year and firm. Specification (1) includes all firm-level controls, plus industry and year fixed effects. Specification (2) adds includes firm fixed effects. All standard errors are clustered at the firm/year level.

regression results to the placebo results, in the specifications without firm-fixed effects, where the baseline results were strongest. All of the placebo results are insignificant, confirming that the changes in volume, drift and volatility are all specific to earnings days. In unreported results, I randomly assign one day for each firm in each quarter to be a placebo earnings day. This alternative placebo test also yields insignificant coefficients on $\Delta Passive$ and $Passive$ in all reduced-form regressions.

F. Addressing Competing Hypotheses

This subsection discusses two alternative explanations for my findings on decreased market efficiency, and its correlation with passive ownership: (1) Regulation Fair Disclosure (Reg FD), which reduced early release of earnings information and (2) the rise of algorithmic trading (AT), which can reduce the returns to informed trading (see e.g. Weller (2017), Farboodi and Veldkamp (2017)). It is not possible to discuss every alternative hypothesis, so outside of explicitly testing these two alternatives, I rely on the quasi-exogenous variation in passive ownership from index addition/rebalancing in the next section to overcome any remaining identification concerns.

In addition to identification concerns, the reduced-form regressions could suffer from omitted variable bias. Most passive ownership is determined by mechanical rules derived from observable signals like market capitalization and past returns. This implies that it may be possible to select

a large set of stock/firm characteristics to explain all of the variation in passive ownership. My results would be biased if these underlying characteristics were driving the changes in pre-earnings price informativeness. I find this unlikely, as a significant amount of the differences in passive ownership across stocks is determined by index membership, which is sticky for some indices, and hard to predict for others. Firms that have been in the S&P 500 index for many years would not necessarily be added to the index today, even if they meet all the criteria for index addition. For other indices like the Russell 1000, there is a sharp size cutoff in the index addition rule[16], which makes it difficult to predict index membership around the cutoff. The difficulty of predicting index membership, and as a result predicting passive ownership, reduces the likelihood that my results are driven by an omitted variables problem.

F.1. Reg FD

Before Reg FD was passed in August, 2000, firms would disclose earnings information to selected analysts before it became public. This information leakage could increase the share of earnings information incorporated into prices before it was formally announced. After Reg FD, firms were no longer allowed selectively disclose material information, and instead must release it to all investors at the same time.

Reg FD could be driving the trends in decreased price informativeness, as there was a large negative shock to information released by firms after it was passed. In Figures 1, 3 and 4, however, all of the information measures continue to trend in the same direction after Reg FD was implemented. Reg FD could still explain these results if the value of the inside information received by analysts before Reg FD decayed slowly. While this is possible, my prior is that information obtained in 2000 would not be relevant for more than a few years. If Reg FD totally explained the decreased pre-earnings informativeness, I would expect the trends in decreased informativeness to level out in the early 2000's. In the data, however, this leveling out does not happen for any of the three measures.

For Reg FD to be driving the reduced-form relationship between passive ownership and pre-earnings price informativeness, it would have to disproportionately affect firms with high passive ownership. This is because all the regressions have year fixed effects, which should account for any level shifts in price informativeness before/after Reg FD was passed. To further rule out this channel, the appendix contains versions of all the reduced-form regressions using only post-2000 data in Tables XVIII, XIX and XX. All of the results are similar using only post-2000 data, suggesting that differences between the pre and post Reg FD eras are not driving my results.

F.2. Rise of Algorithmic Trading Activity

Weller (2017) shows that Algorithmic Trading (AT) activity is negatively correlated with pre-earnings price informativeness. His proposed mechanism is ATs front-run informed traders, re-

[16]There was a sharp size cutoff before the rule change in 2006, see e.g. Wei and Young (2017).

ducing the returns to gathering firm-specific fundamental information. AT activity increased significantly over my sample period, and could be responsible for some of the observed decrease in pre-earnings price informativeness.

It is difficult to measure the role of ATs in the trends toward decreased pre-earnings price informativeness as I cannot directly observe AT activity, and only have reasonable AT activity measures between 2012-2017. I can, however, measure the effect of AT activity on the reduced-form results. For AT activity to influence the regression estimates, it would have to be correlated with passive ownership, which I find plausible because: (1) Passive ownership is higher in large, liquid stocks, where most AT activity occurs. This should not affect my results, however, as I condition on firm size in all the reduced-form regressions (2) High ETF ownership will attract ATs implementing ETF arbitrage. The effect of time trends in AT activity should be absorbed by the year fixed effects.

To rule out this channel, I construct the 4 measures of AT activity used in Weller (2017) from the SEC MIDAS data. MIDAS has daily data for all stocks traded on 13 national exchanges from 2012 to 2017. The AT measures are (1) odd lot ratio, (2) trade-to-order ratio, (3) cancel-to-trade ratio and (4) average trade size. Measures 1 and 3 are positively correlated with AT activity, while the opposite is true for measures 2 and 4. Consistent with Weller (2017), I (1) Winsorize each of the AT activity variables at the 1% and 99% level by year to minimize the effect of reporting errors, (2) calculate a moving average for each of these measures in the 21 days leading up to each earnings announcement, and (3) take logs to reduce heavy right-skewness. Only 1% of MIDAS data cannot be matched to CRSP, so the 87% drop in sample size relative to previous regressions is coming almost entirely from the year restrictions.

I run the following modified versions of the reduced-form regressions:

$$\begin{aligned}\Delta_{(t,t-n)}CAV_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + \gamma X_{i,t-n} + \\ \phi\Delta_{(t,t-n)}ATActivity_{i,t} + \text{Fixed Effects} + e_{i,t}\end{aligned} \tag{12}$$

$$\begin{aligned}DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=2}^{10}\phi_j \mathbf{1}_{SUEdecile=j} + \gamma X_{i,t-1} + \\ \psi ATActivity_{i,t} + \text{Fixed Effects} + e_{i,t}\end{aligned} \tag{13}$$

$$\begin{aligned}\Delta_{(t,t-n)}QVS_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + \gamma X_{i,t-n} + \\ \phi\Delta_{(t,t-n)}ATActivity_{i,t} + \text{Fixed Effects} + e_{i,t}\end{aligned} \tag{14}$$

In regression 12 $\Delta_{(t,t-n)}ATActivity_{i,t}$ is a vector of year-over-year changes in the 4 AT activity measures. In regression 13 $ATActivity_{i,t}$ is a vector containing the levels of the 4 AT activity measures. Regression 14 is run with annual data, so I first calculate average $ATActivity_{i,t}$ across the 4 earnings announcements each year, and then calculate year-over-year changes. As a baseline,

I re-run the previous regressions on the sub-sample matched to the MIDAS data – these regressions are labeled "Baseline" in the corresponding tables. The regressions with all the AT measures included are labeled "+ AT Controls". Given the limited time-series available for the AT measures, I only run the one-year and three-year difference specifications.

Tables XV, XVI and XVII contain the results. Only the results for the pre-earnings drift and the earnings day share of volatility are significant in the matched subsample. For the specifications that are significant in the subset of my original sample that can be matched to the MIDAS data, adding the AT activity controls does reduce the coefficient on passive ownership/change in passive ownership, but the sign and statistical significance is unchanged. This implies that increased AT activity may partially explain the observed decrease in market efficiency, but increasing passive ownership is still an important factor in decreased pre-earnings price informativeness.

G. Evaluating Predictions from Theory

In the model from Section II, the effect of increasing passive ownership on price informativeness is ambiguous. The null hypothesis that passive ownership is too small to have an observable effect is rejected in the data. We can also reject the second alternative hypothesis that prices became more informative. The reduced-form results suggest that the decrease in informed agents is large enough to decrease market efficiency, at least for firm-specific information, and this dominates the secondary channel of more effective hedging.

In the next section, I present direct evidence that less firm-specific information is being gathered for stocks with high passive ownership, which is consistent with a decrease in the number of informed investors. This does not, however, explain *why* less firm-specific information is being gathered.

The number of informed agents could be endogenized in the model by allowing agents to pay a fixed cost to learn the private signal, and decide whether they want to be informed or uninformed. Suppose that, for reasons outside the model, there is a significant decrease in the number of informed agents. Empirically, this might have occurred because active managers consistently under-performed passive strategies for several years, and investors switched their money into index funds. The resulting decrease in price informativeness should entice more agents to become informed, as the expected return to acquiring the private signal would be high. Eventually, enough uninformed agents would become informed that price informativeness would return to the level it was before the shock. Based on the empirical results, however, something outside the model is preventing this from occurring.

One possible explanation is the rise of algorithmic trading. Weller (2017) shows that algorithmic trading reduces the returns to gathering firm-specific information. With the growth of algorithmic trading, fewer uninformed agents are willing to pay to acquire the signal after the shock to the number of informed agents, even though the expected distance between the price and the fundamental value is large.

Another potential explanation is that the costs of trading baskets of stocks, relative to individual equities, has changed. To capture this, the model could be enriched further, allowing investors to

pay a fixed cost to learn about a particular stock, or about a basket of stocks. If the relative cost of trading the basket decreases, which empirically may be the result of liquid ETFs and low ETF fees, it may be more profitable to acquire the signal about the basket than the individual stock. Glosten et al. (2016) shows that ETFs have increased the systematic information content of stock prices. This is consistent with investors acquiring signals about baskets of stocks, and using ETFs to trade on these signals.

IV. Exploring Potential Mechanisms

Section III shows a negative relationship between passive ownership and price informativeness. In the context of the model, this could be explained by a decrease in the number of informed agents. In this section, I present reduced-form evidence consistent with fewer investors gathering firm-specific information.

A. Analyst Attention

One potential explanation for passive ownership decreasing price informativeness is that passive managers, as well as investors in passive funds, lack strong incentives to gather and consume firm-specific information. Passive funds trade on mechanical rules, such as S&P 500 index membership (SPY), or the 100 lowest volatility stocks in the S&P 500 (SPLV). Given that these trading strategies are implemented on public signals, they do not require accurate private forecasts of firm fundamentals. As a stock becomes more mispriced, however, the return to gathering fundamental information increases, so it is not obvious which effect will dominate in equilibrium. To test this hypothesis, I regress levels, and changes in analyst coverage/accuracy on passive ownership:

$$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{15}$$

Controls in $X_{i,t}$ include market capitalization and institutional ownership. Fixed effects include industry, year and firm.

$$\Delta_{(t,t-5)} Outcome_{i,t} = \alpha + \beta \Delta_{(t,t-5)} Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{16}$$

$\Delta_{(t,t-5)}$ is the change from year t to year $t-5$. Controls in $X_{i,t}$ include institutional ownership, lagged inst. ownership, market cap, lagged market cap. Fixed effects include industry, year and firm.

In regressions 15 and 16, the outcomes of interest are (1) the number of analysts covering a stock, (2) the absolute distance between the consensus forecast and the realized earnings, divided by the absolute value of the consensus forecast, which I will call *accuracy* and (3) the probability that the consensus forecast has the same sign as the realized earnings. For the accuracy regressions, I exclude firms with a consensus forecast of 1 cent or less (in absolute value) to minimize the effect of outliers. Accuracy is then Winsorized at the 1% and 99% level each year.

	# Analyst Estimates		$\frac{\lvert realized-consensus \rvert}{\lvert consensus \rvert}$		P(consensus correct sign)	
	(1)	(2)	(1)	(2)	(1)	(2)
Passive Ownership	-1.307***		0.298*		-0.171***	
	(0.437)		(0.166)		(0.028)	
Increase in Passive Ownership		-5.693***		0.941***		-0.198***
		(0.704)		(0.252)		(0.043)
Unconditional Avg.	6.28	6.28	0.39	0.39	0.95	0.95
Observations	155,983	82,413	126,685	65,338	127,991	66,308
R-Squared	0.151	0.146	0.012	0.014	0.006	0.007
Industry FE	YES	YES	YES	YES	YES	YES
Year FE	YES	YES	YES	YES	YES	YES
Firm FE	YES	YES	YES	YES	YES	YES
Firm-Level Controls	YES	YES	YES	YES	YES	YES

Table IV Analyst Coverage/Accuracy and Passive Ownership. This table contains the results for two sets of regressions:

$$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Controls in $X_{i,t}$ include market capitalization and institutional ownership. Fixed effects include industry, year and firm.

$$\Delta_{(t,t-5)} Outcome_{i,t} = \alpha + \beta \Delta_{(t,t-5)} Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Controls in $X_{i,t}$ include institutional ownership, lagged institutional ownership, market capitalization, lagged market capitalization. Fixed effects include industry, year and firm.

The sample is all annual earnings announcements. To determine the consensus forecast, I take the equal-weighted average of all analyst forecasts on the last statistical period in IBES before earnings are released.

Table IV contains the regression results. Consistent with decreased information gathering, high levels and increases in passive ownership are correlated with the fewer analysts covering a stock, lower analyst accuracy, and a lower likelihood of the consensus forecast matching the sign of realized earnings.

B. Liquidity

A more mechanical explanation for the negative relationship between passive ownership and pre-earnings price informativeness is that high passive ownership decreases the number of shares available for trading (float), which leads to higher price impact and transaction costs. Section B of the Appendix shows that increases in passive ownership are correlated with increased average liquidity, but decreased liquidity around earnings announcements. This result, however is not robust using only quasi-exogenous increases in passive ownership arising from S&P 500 index addition and Russell 1000/2000 reconstitution.

C. *Downloads of SEC Filings*

One measure of investor attention is the number of downloads of SEC filings (see e.g. Loughran and McDonald (2017)). If passive managers, and investors in passive funds, do not gather fundamental information, the number of downloads of SEC filings might be lower for firms with high passive ownership. To test this, I run the following regression:

$$\Delta_{(t-1,t)} DL_{i,t} = \alpha + \beta \Delta_{(t-1,t)} Passive_{i,t} + X_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{17}$$

$\Delta_{(t-1,t)}$ is the year-over-year change. $DL_{i,t}$ is the number of non-robot downloads of 10-K's, 10-Q's and 8-K's in the 30 days before earnings announcements. Robot downloads include web crawlers, index page requests and individual IPs with large number of downloads in a single day. This definition is based on data made available by Bill McDonald, originally derived from the Edgar Server Log between 2003 and 2015. I exclude robot downloads as they may automatically download all filings at release, or update a database periodically for reasons other than information gathering. Controls in $X_{i,t}$ include size, idiosyncratic volatility, institutional ownership and passive ownership. Fixed effects include year, day of the week and firm. Over time, the average number of downloads has been increasing.

Table V contains the regression results. Consistent with decreased information gathering, firms with increases in passive ownership experience decreases in pre-earnings downloads of SEC filings. The increase in downloads on the earnings days themselves has many possible explanations, including the increased earnings day volatility for firms with high passive ownership, as shown in Section III.

D. *Response to Earnings News*

Buffa et al. (2014) propose a model where stocks with a higher share of "buy and hold" investors are more responsive to cash flow news. In the model, buy and hold investors distort prices, so informed investors underweight these stocks. When the good cashflow news arrives, the informed investors were previously underweight these stocks, so their diversification motive is weak, and they buy. In relating this model to my empirical setup, I treat buy and hold investors as passive owners and the cashflow news as earnings announcements.

To test the model's predictions, I run the following regression:

$$\begin{aligned} r_{i,t} = \alpha + \beta_1 SUE_{i,t} + \beta_2 1_{SUE<0} + \gamma_1 \left(SUE_{i,t} \times Passive_{i,t}\right) + \\ \gamma_2 \left(1_{SUE<0} \times Passive_{i,t}\right) + \phi X_{i,t} + \text{Fixed Effects} + e_{i,t} \end{aligned} \tag{18}$$

Here, $r_{i,t}$ denotes the market-adjusted return on the effective quarterly earnings date. SUE is defined as in Novy-Marx (2015): $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$.[17] Controls in $X_{i,t}$ include lagged firm size, lagged idiosyncratic volatility over the past year, and institutional ownership. Fixed

[17] Results are similar when calculating SUE relative to IBES estimates using the method in Anson, Chambers, Black, Kazemi, Association, et al. (2012).

	Change in # of non-Robot Downloads	
	30 Days Before ED	Earnings Day
Change in Passive	-0.533***	0.378**
	(0.072)	(0.167)
Observations	119,712	119,807
R-Squared	0.076	0.049
Firm-Level Controls	YES	YES
Year/Industry FE	YES	YES
Firm FE	YES	YES

Table V Downloads and Passive Ownership. This table contains the results of the following regression:

$$\Delta DL_{i,t} = \alpha + \beta \Delta Passive_{i,t} + X_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Controls in $X_{i,t}$ include size, idiosyncratic volatility, institutional ownership and passive ownership. Fixed effects include year, day of the week and firm. Over time, the average number of downloads has been increasing.

effects include year, 2-digit SIC industry and firm. The coefficients of interest are γ_1 and γ_2, which capture the effect of passive ownership on responses to earnings surprises, and the asymmetric effect of responses to negative earnings surprises.

Table VI contains the regression results. Consistent with the models, firms with a high share of passive ownership are more responsive to earnings news.

V. Robustness to Quasi-Exogenous Variation in Passive Ownership

In this section, I exploit S&P 500 index additions, as well as Russell 1000/2000 reconstitutions to create a causal link between increases in passive ownership and decreases in pre-earnings price informativeness.

A. S&P 500 Index Addition

Each year, a committee from Standard & Poor's selects firms to be added/removed from the S&P 500 index. For a firm to be added to the index, it has to meet criteria set out by S&P, including a sufficiently large market capitalization, a specific industry classification and financial health. Once a firm is added to the S&P 500 index, it experiences a large increase in passive ownership, as many index funds and ETFs buy the stock.

	Market-Adjusted ED Return	
	(2)	(3)
SUE	0.181***	0.210***
	(0.011)	(0.012)
$1_{SUE<0}$	-0.863***	-0.726***
	(0.033)	(0.039)
SUE x Passive Share	3.270***	2.577***
	(0.219)	(0.227)
$1_{SUE<0} \times PassiveShare$	-0.836	-3.299***
	(0.524)	(0.608)
Observations	422,569	422,569
Firm Controls	Yes	Yes
Year FE	Yes	Yes
Industry FE	Yes	Yes
Firm FE	No	Yes

Table VI Passive Ownership and Response to Earnings News. This table contains the results of the following regression:

$$r_{i,t} = \alpha + \beta_1 SUE_{i,t} + \beta_2 1_{SUE<0} + \gamma_1 (SUE_{i,t} \times Passive_{i,t}) + \gamma_2 (1_{SUE<0} \times Passive_{i,t}) + \phi X_{i,t} + \text{Fixed Effects} + e_{i,t}$$

$r_{i,t}$ is the market-adjusted return on the quarterly effective earnings date. Controls in $X_{i,t}$ include lagged firm size, lagged idiosyncratic volatility over the past year, and institutional ownership. Fixed effects include year, 2-digit SIC industry and firm.

I obtain daily S&P 500 index constituents from Compustat. Motivated by the size, industry and financial health criteria, I select a group of control firms that reasonably could have been added to the index at the same time as the treated firms. One year before index addition, I sort firms into two-digit SIC industries. Then, within each industry, I sort firms into quintiles based on market capitalization (size) and growth rate of market capitalization over the past year. The control group is all firms in the same industry/size/growth-rate bucket that were not added to the index over the next two years.[18] For example, Sun Microsystems was added to the Index in August, 1992. while one of the corresponding control firms, Seagate Technologies, was not added to the Index until August, 1996.

For the reduced-form regressions in differences (Pre-Earnings Volume and Earnings Day Share of Volatility) I run the following regression:

$$\Delta_{(t-1,t+1)} Outcome_{i,t} = \alpha + \gamma Added_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{19}$$

[18] It is also possible to sort on three-digit SIC industries, but this leaves many firms without a control firm as not every SIC-3 industry has at least 2 firms in each of the 25 size/growth rate buckets.

where $\Delta_{(t-1,t+1)}$ is the change from the year before index addition to the year after index addition.[19]

For the Pre-Earnings Volume regression, there are 4 observations per firm – each is the change from the same fiscal quarter in the year before addition to the year after addition. For the Earnings Day Volatility regression, there is one observation per firm. $Added_{i,t}$ is an indicator variable equal to one if the firm was added to the S&P 500 index. The coefficient of interest is γ, the treatment effect of being added to the index.

For the Pre-Earnings drift regression, the reduced-form was in levels, so I run the following Difference in Differences specification:

$$Outcome_{i,t} = \alpha + \beta Added_{i,t} + \tau After_{i,t} + \gamma Added_{i,t} \times After_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{20}$$

In this regression, there are 8 observations for each firm: 4 in the year before index addition, and 4 in the year after index addition. $Added_{i,t}$ is an indicator variable equal to one if the firm was added to the S&P 500 index. $After_{i,t}$ is an indicator equal to one for the observations after the firm was added. The coefficient of interest is the interaction term, γ.

For all three regressions, the fixed effects include 2-digit SIC industry, size quintile, growth rate quintile and year/month of index addition. With these controls, the regressions are only exploiting variation between the treatment and control firms at the same point in time.

One concern is that because selection is by committee, the increase in passive ownership is not fully exogenous to firm fundamentals. Partially alleviating this concern is that, according to S&P (2017): "Stocks are added to make the index representative of the U.S. economy, and is not related to firm fundamentals." As an additional check, in the next subsection I focus on Russell 1000/2000 reconstitution, which is based on a mechanical rule.

To test the similarity of the treated and control groups, Figure 6 shows the levels and changes in passive ownership for the control firms and treatment firms around the time of index addition. Both groups of firms have similar levels and pre-addition changes in passive ownership.

Table VII contains the regression results. For comparison, I included a row with the reduced form estimates, which correspond to the 1-year changes specification with firm fixed effects estimated in Section 3. The average year-over-year increase in passive ownership for a firm added to the S&P 500 index is 2.2%, so the implied elasticity is the coefficient of interest, γ, divided by 0.022. For all three specifications, the results have the same sign and statistical significance as the reduced-form regressions. The implied elasticities are substantially larger than the reduced-form results, but my prior is that this is because 2.2% understates the true increase in passive ownership associated with index addition: There are many institutional investors which do not show up in the Thompson S12 data which track the S&P 500 index and buy these stocks after they are added.

[19]The year of index addition is omitted for two reasons: (1) The increase in passive ownership associated with index addition does not all occur immediately, but rather over the three quarters after index addition (2) There are other effects associated with S&P 500 index addition, including an increase in relative valuation (see e.g. Morck and Yang (2001)) and an increase in media coverage (Engelberg and Gao (2011)). Skipping a year gives time for these index addition effects to die out. In unreported results, I find that not skipping a year yields similar point estimates, but larger standard errors.

	Pre-Earnings Volume Differences	Pre-Earnings Drift Levels	Earnings Day Share of QV Differences
Added	-0.0537**		0.0116**
	(0.026)		(0.006)
Interaction		-0.0904*	
		(0.051)	
R-squared	0.083	0.073	0.129
Reduced Form	-1.090***	-0.436**	0.0551*
First Stage	2.20%	2.20%	2.20%
Implied Elasticity	-2.44	-4.11	0.53
Year/Month FE	YES	YES	YES
Industry FE	YES	YES	YES
Size Quintile FE	YES	YES	YES
Growth Quintile FE	YES	YES	YES
Treated Firms	614	614	614
Control Firms	1,479	1,479	1,479

Table VII Effects of S&P 500 Index Addition. For Pre-Earnings Volume and Earnings Day Share of QV:

$$\Delta_{(t-1,t+1)} Outcome_{i,t} = \alpha + \gamma Added_{i,t} + \text{Fixed Effects} + e_{i,t}$$

For Pre-Earnings Drift:

$$Outcome_{i,t} = \alpha + \beta Added_{i,t} + \tau After_{i,t} + \gamma Added_{i,t} \times After_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Added firms are those which were added to the S&P 500 index. Control firms are in the same industry, same size and same growth rate quintile as the control firms. Fixed effects include industry, size quintile, growth rate quintile, and year/month of index addition.

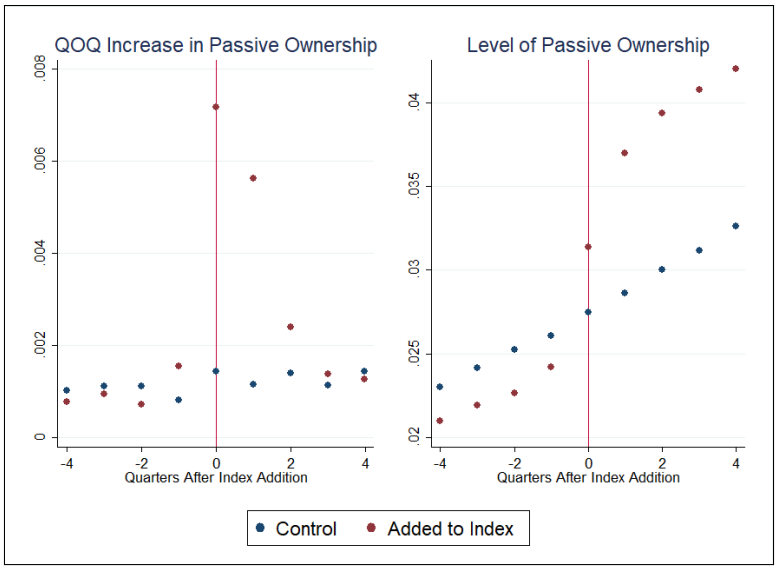

Figure 6. S&P 500 Index Addition and Changes in Passive Ownership. Average level and increase in passive ownership for control firms and firms added to the S&P 500. Control firms are all firms in the same 2-digit SIC industry, in the same size and growth rate quintiles that were not added to the S&P 500 index over the next two years.

A natural extension is to examine firms which were dropped from the S&P 500 index, which experience a decrease in passive ownership. This is a less ideal experiment than index addition, as firms are usually dropped from the index for poor performance or lack of liquidity, which is related to firm fundamentals. Section H of the Appendix has more details on the effect of index deletion.

B. Russell 1000/2000 Index Reconstitution

The Russell 3000 contains approximately the 3000 largest stocks in the United States stock market. Each May, FTSE Russell selects the 1000 largest stocks by float to be members of the Russell 1000, while it selects the next 2000 largest stocks by float to be members of the Russell 2000.[20] Both of these indices are value-weighted, so moving from the 1000 to the 2000 significantly increases the share of passive ownership in a stock. The firm goes from being the smallest firm in an index of large firms, to the biggest firm in an index of small firms, increasing its relative weight by a factor of 10 (see e.g. Appel, Gormley, and Keim (2016)).

The increase in passive ownership corresponding to S&P 500 index addition is not a perfect natural experiment because firms are not added at random, added firms receive increased attention, and added firms may start marketing their stock differently to institutional investors. The increase in passive ownership associated with the Russell reconstitution sidesteps many of these issues, as moving from the 1000 to the 2000 is based on a mechanical rule, rather than committee selection.

[20]This rule changed in 2006 – to reduce turnover between the two indices, Russell now has a bandwidth rule: As long as the firm's market capitalization is within 5% of the 1000th ranked stock, it will remain in the same index it was in the previous year. Given that this is still a mechanical rule, however, the increases in passive ownership are still plausibly exogenous to firm fundamentals.

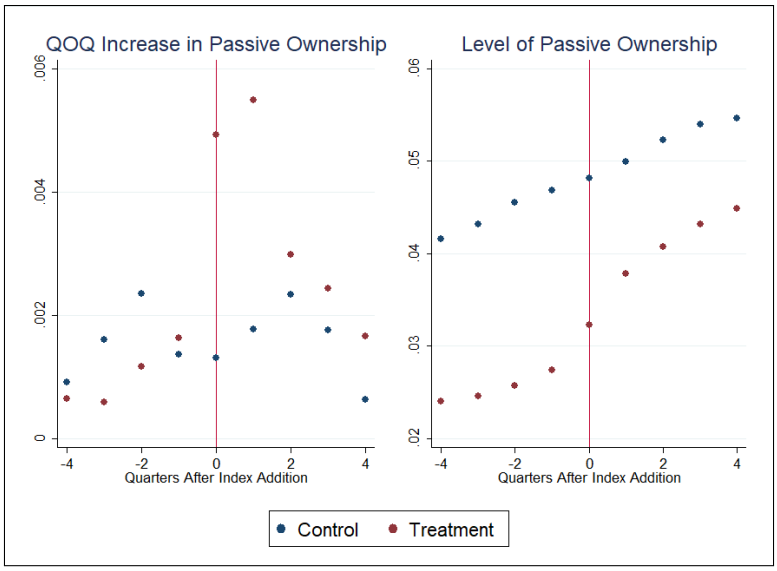

Figure 7. Russell 1000/2000 Reconstitution and Changes in Passive Ownership. Average level and increase in passive ownership for control firms and firms moved from the Russell 1000 to the Russell 2000. Control firms are all firms in the Russell 3000 ranked 900 to 1100 that did not move from the 1000 to the 2000 or from the 2000 to the 1000.

Further, because the firm's market capitalization shrunk, it is less likely to change the way the firm is marketing itself to institutions.

I obtain Russell 1000/2000 membership between 1996 and 2012 from the Wei and Young (2017) replication files. The treated firms are those that switched from the Russell 1000 to the Russell 2000. The control firms are the other firms with June ranks between 900 and 1100 that did not switch between the 1000 and the 2000 or between the 2000 and the 1000. Results are similar when restricting the control firms to those that stayed in the Russell 1000. This classification involves a look-ahead bias, as I am using the ex-post changes in membership to identify changes in passive ownership.

Figure 7 compares the level, and the increase in passive ownership around the index rebalancing date. While the pre-addition changes are similar, the levels are different – this is driven by the firms ranked 1001-1100 having a higher average level of passive ownership because they are the largest firms in a value-weighted index of small firms.

I re-run regressions 19 (pre-earnings volume and earnings day volatility) and 20 (pre-earnings drift) with the Russell 1000/2000 index reconstitutions, and a slightly different set of controls and timing assumptions. I remove the industry, size quintile and growth rate quintile fixed effects, as I am not using these to select control firms. As for timing, I am comparing the year before index reconstitution, ending in April, and the year following reconstitution, starting in August. This is because the rankings are determined in May, so investors may trade in advance of the actual rebalancing in June. Further, the rankings are usually released at the end of June, but sometimes they are released in early July. July is excluded to avoid any of the trading associated with index

rebalancing influencing the regression results.[21]

Table VIII contains the regression results. For comparison, I included a row with the reduced form estimates, which correspond to the 1-year changes specification with firm fixed effects estimated in Section 3. The average increase in passive ownership from May to August for a firm moving from the Russell 1000 to the 2000 is 1.7%, so the implied elasticity is the coefficient of interest, γ, divided by 0.017.

For pre-earnings volume and drift, the results have the same sign and statistical significance as the reduced-form regressions. The results for earnings day volatility have the opposite sign and are insignificant. Part of this could be due to the volatility of QVS. Given the relatively short sample period (1996-2012), and the smaller number of treated and control firms, this estimate is likely noisy.

As with the S&P 500 results, the implied elasticities are substantially larger than the reduced-form estimates, but I believe 1.7% understates the true increase in passive ownership associated with index addition: There are many institutional investors which track the Russell indices which do not show up in the Thompson S12 data.

A natural extension is to look at the firms which experience a decrease in passive ownership when they move from the Russell 2000 to the Russell 1000. In Section H I show that this treatment effect is washed out by the time trend toward increased passive ownership.[22]

[21] Unlike the S&P 500, where firms remain in the index for long stretches of time, the Russell indices are rebalanced annually, so one year after moving from the 1000 to the 2000, the firm may switch back. To avoid picking up the effects of firms switching back and forth, I do not skip a year after index additions.

[22] Another plausibly exogenous change in passive ownership arises when firms move from outside the Russell 3000 to inside the Russell 3000, which results in an increase in passive ownership. While this is potentially interesting, there are sample selection issues, as these micro caps often fail to appear in IBES or Compustat.

	Pre-Earnings Volume Differences	Pre-Earnings Drift Levels	ED Share of QV Differences
Treated	-0.0559***		-0.00237
	(0.018)		(0.006)
Interaction		-0.100**	
		(0.043)	
R-squared	0.005	0.006	0.030
Reduced Form	-1.090***	-0.436**	0.0551*
First Stage	1.70%	1.70%	1.70%
Implied Elasticity	-3.288	-5.882	-0.139
Year/Month FE	YES	YES	YES
Treated Firms	479	479	479
Control Firms	1,284	1,284	1,284

Table VIII Effects of Russell 1000/2000 Index Reconstitution. For Pre-Earnings Volume and Earnings Day Share of QV:

$$\Delta_{(t-1,t+1)} Outcome_{i,t} = \alpha + \gamma Moved_{i,t} + \text{Fixed Effects} + e_{i,t}$$

For Pre-Earnings Drift:

$$Outcome_{i,t} = \alpha + \beta Moved_{i,t} + \tau After_{i,t} + \gamma Moved_{i,t} \times After_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Added firms are those which were moved from the Russell 1000 to the Russell 2000. Control firms are those that were ranked 900-1100 by Russell, but did not switch from the 1000 to the 2000 or from the 2000 to the 1000. Fixed effects include year/month of index rebalancing.

VI. Conclusion

Increases in passive ownership have lead to decreased pre-earnings-announcement price informativeness. When passive ownership in a stock increases, there is less pre-earnings trading, a smaller pre-earnings drift and a larger share of volatility on earnings days. These results are robust to only exploiting quasi exogenous variation in passive ownership that arises from index addition and rebalancing.

One potential mechanism is that passive managers, as well as investors in passive funds, lack strong incentives to gather and consume firm-specific information. Consistent with this channel, firms with increases in passive ownership experience decreases in the number of analysts covering the stock, decreases in the accuracy of the remaining analysts, and fewer downloads of SEC filings.

Relative to total institutional ownership, passive ownership is still relatively small, owning around 10% of total US market capitalization. The model from Section II predicts a convex relationship between decreases in informed investors and price informativeness, so further increases in passive ownership could lead to even larger declines in efficiency.[23] It will be interesting to re-measure these efficiency trends in the future, and see if these non-linear effects kick in.

[23] See Section J of the Appendix for details

Appendix A. Data Details

Daily Volume (from CRSP): Number of shares traded across all US exchanges. This quantity is not adjusted for splits during the month and it does not contain over-allotments. Beginning in November 2008, volume also includes trades on the BATS Exchange, which now accounts for over 10% of all US equity trading.

I/B/E/S: Before 1998, nearly 90% of observations in IBES have an announcement time of "00:00:00", which implies the release time is missing. In 1998 this share drops to 23%, further drops to 2% in 1999, and continues to trend down to 0% by 2015. This implies that before 1998, if the earnings release date was a trading day, I will always classify that as the effective earnings date, even if earnings were released after markets closed, and it was not possible to trade on that information until the next trading day.

This time-variation in missing observations is not driving my results for two reasons: (1) I re-run every regression using only post-2000 data and the results are similar (2) For the pre-earnings drift, and pre-earnings volume, I am measuring returns/volume leading up to an earnings announcement. These missing earnings times could only move the effective earnings date earlier in time, which would bias both of my measures toward finding nothing. If volume dropped significantly on the last trading day before the earnings announcement, this would not be included in my pre-earnings volume measure for observations with a missing announcement time. For the pre-earnings drift, and the earnings day share of volatility, it would lead to selecting days where no news was released, which likely have smaller, rather than larger moves on average, pushing DM toward 1, and QVS toward 1.6%.

Appendix B. Liquidity

A mechanical explanation for the relationship between increased passive ownership and large (absolute) returns on earnings days is that high passive ownership decreases the number of shares available for trading (float), which leads to higher price impact and transaction costs.

I examine the effect of increases in passive ownership on liquidity, as measured by the bid-ask spread. I calculate the daily bid-ask spread for firm i at time t as: $BA_{i,t} = (ask_{i,t} - bid_{i,t})/ask_{i,t}$ using the closing *bid* and *ask* in CRSP data. All results in this section are similar using the definition of bid-ask spread from Abdi and Ranaldo (2017). Figure 8 shows that the spread expands the day before an earnings announcement, and quickly reverts to pre-announcement levels.

To understand the effect of passive ownership on average liquidity, I run the following regression:

$$\Delta_{(t,t-n)}\overline{BA}_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t} \tag{B1}$$

$\overline{BA}_{i,t}$ is the average bid ask spread across all days in year t. Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC

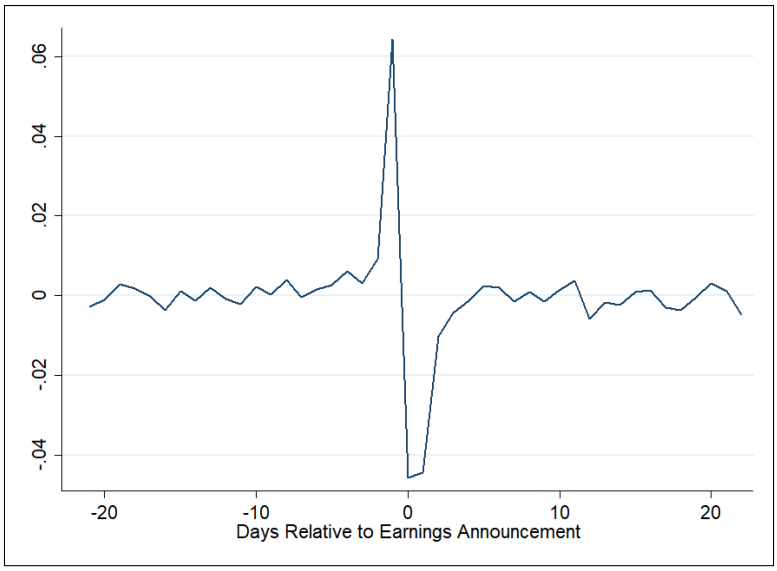

Figure 8. Change in the Bid Ask Spread Around Earnings Announcements. Average daily change in bid-ask spread. Computed as (ask-bid)/ask, using daily data in CRSP.

industry, year and firm.

The results are in Table IX. Across all specifications, increases in passive ownership are negatively correlated with the average bid-ask spread. This suggests that stocks become more liquid as passive ownership increases, consistent with results in Glosten et al. (2016). It also implies that decreases in float arising from increases in passive ownership are not driving my results.

As shown in Figure 8, liquidity dries up before earnings announcements. To understand the relationship between passive ownership and pre-earnings liquidity, I run the following regression:

$$\Delta_{(t,t-n)} BA_{i,(\tau-2,\tau-1)} = \alpha + \beta \Delta_{(t,t-n)} Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t} \tag{B2}$$

$BA_{i,(\tau-2,\tau-1)}$ is the change in the bid-ask spread between $\tau - 2$ and $\tau - 1$, where τ is an earnings announcement date. $\Delta_{(t,t-n)}$ is the change from calendar year $t - n$ to calendar year t. I only look at year-over-year changes to avoid differences in bid-ask spreads before annual earnings announcements and quarterly announcements or seasonal effects. Controls in $X_{i,t-n}$ include lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership, calculated as the sum of holdings in the 13-F filings. I also condition on the growth in market capitalization from $t - n$ to t.

The results are in Table X. Although not all specifications are significant, there is a positive relationship between increases in passive ownership, and increases in the bid-ask spread before earnings announcement. This is consistent with the results in the main body of the paper, where increased passive ownership leads to less informative prices, and more adverse selection before earnings announcements.

When using the S&P 500 addition setup, the results have the same sign as the reduced form

		1-year		3-year		5-year	
		(1)	(2)	(1)	(2)	(1)	(2)
Increase in Passive Ownership	1-year	-0.0473***	-0.0415***				
		(0.011)	(0.001)				
	3-year			-0.0680***	-0.0460***		
				(0.016)	(0.002)		
	5-year					-0.0885***	-0.0492***
						(0.017)	(0.002)
Observations		127,270	127,270	99,109	99,109	77,920	77,920
R-Squared		0.277	0.296	0.357	0.407	0.384	0.454
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

Table IX Passive Ownership and Bid Ask Spread:Annual Averages. Estimates of β from:

$$\Delta_{(t,t-n)}\overline{BA}_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$\overline{BA}_{i,t}$ is the average bid ask spread across all days in year t. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Change in passive ownership is expressed as a decimal, so $0.01 = 1\%$ increase. Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry, year and firm. Specification (1) includes all firm-level controls, plus industry and year fixed effects. Specification (2) adds includes firm fixed effects. All standard errors are clustered at the firm/year level.

		1-year		3-year		5-year	
		(1)	(2)	(1)	(2)	(1)	(2)
Increase in Passive Ownership	1-year	0.0149***	0.011				
		(0.006)	(0.009)				
	3-year			0.00134	0.00211		
				(0.006)	(0.007)		
	5-year					0.0156***	0.0149**
						(0.005)	(0.007)
Observations		92,144	92,144	77,246	77,246	65,183	65,183
R-Squared		0.020	0.010	0.022	0.011	0.025	0.012
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

Table X Passive Ownership and Bid Ask Spreads:Earnings Days. Estimates of β from:

$$\Delta_{(t,t-n)}BA_{i,(\tau-2,\tau-1)} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$BA_{i,(\tau-2,\tau-1)}$ as the change in the bid-ask spread between $\tau-2$ and $\tau-1$, where τ is an earnings announcement date. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Change in passive ownership is expressed as a decimal, so $0.01 = 1\%$ increase. Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry, year and firm. Specification (1) includes all firm-level controls, plus industry and year fixed effects. Specification (2) adds includes firm fixed effects. All standard errors are clustered at the firm/year level.

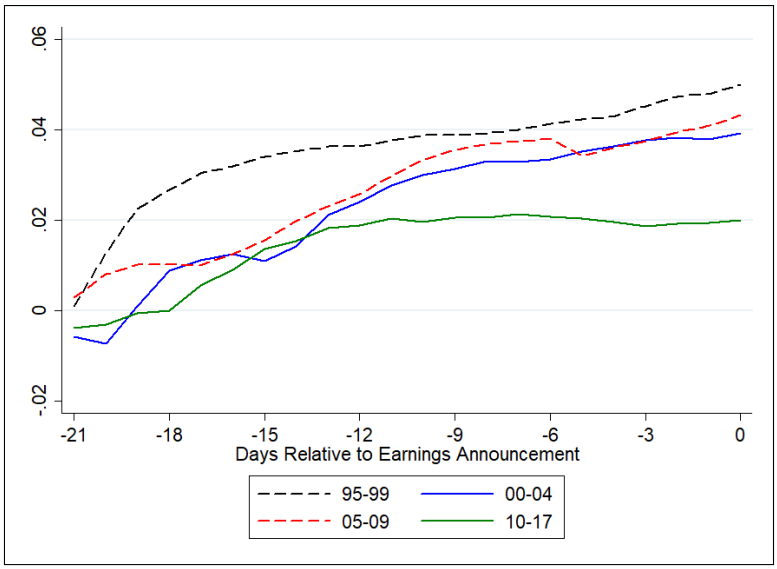

Figure 9. Placebo Test: Pre-Earnings Volume. Plot of $\beta_{\{j==-\tau\}}$ estimated from the regression:

$$\overline{CAV_{i,j,t}} = \alpha + \sum_{\tau=-21}^{0} \beta_\tau \mathbf{1}_{\{j==-\tau\}} + \text{Fixed Effects} + e_{i,j,t}$$

Where t denotes a placebo earnings date. Placebo earnings dates are randomly assigned within each quarter for each firm.

results, but are insignificant. When using the Russell reconstitution, the results are either insignificant or go the wrong way.

Appendix C. Trend Placebo Tests

This section replicates Figures 1 (decrease in pre-earnings volume), 2 (decrease in pre-earnings drift) and 4 (increase in earnings day volatility), except replaces the true earnings dates with a randomly selected date for each firm each quarter. In all three cases, there is no trend toward decreased informativeness on the placebo earnings dates.

Appendix D. Additional Pre-Earnings Volume Results

Rather than look at the 22 days before an earnings announcement, I expand the analysis to trading 60 days before the earnings announcement. This roughly corresponds to the time of the last earnings announcement. A concern with the regression specification in Equation 3 (the regression of cumulative abnormal volume on days-before-earnings-announcement indicator variables) is that average earnings day volume has increased, so the relative volume on the days leading up to the earnings days would appear to *mechanically* decrease in a regression with year fixed effects. Figure 12 shows the cross-sectional median pre-earnings volume, which exhibits the same decline in pre-

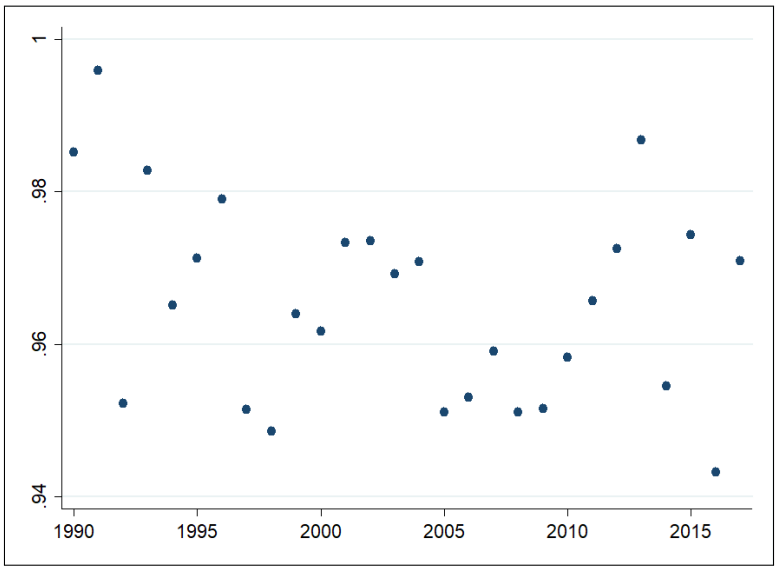

Figure 10. Placebo Test: Pre-Earnings Drift. This figure plots the cross-sectional average of the drift magnitude measure, $DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, by year where t denotes a placebo earnings date. Placebo earnings dates are randomly assigned within each quarter for each firm.

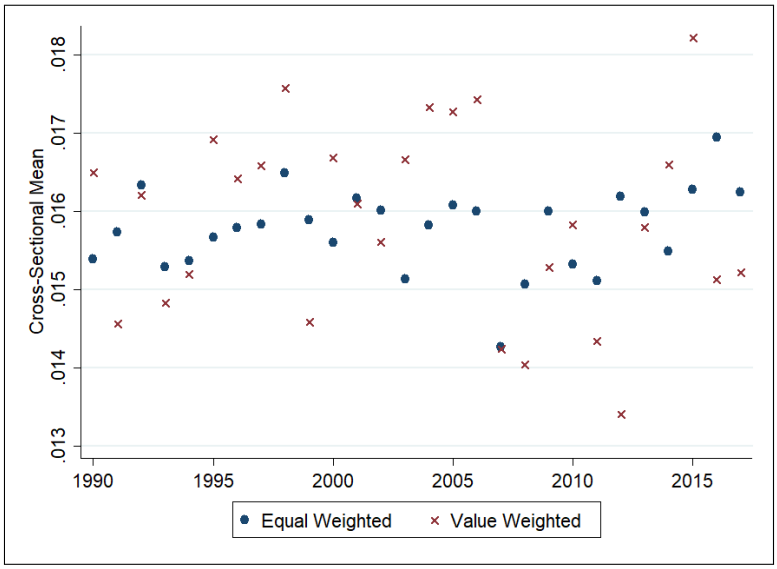

Figure 11. Placebo Test: Earnings Day Volatility. This figure plots the share of market-adjusted quadratic variation occurring on placebo earnings days. For firm i in year t the quadratic variation share (QVS) is defined as: $QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, where r denotes a market-adjusted daily return. The numerator is the sum of squared returns on the 4 placebo earnings dates, while the denominator is the sum of squared returns on all trading days in calendar year t. Placebo earnings dates are randomly assigned within each quarter for each firm.

earnings volume as Figure 1.

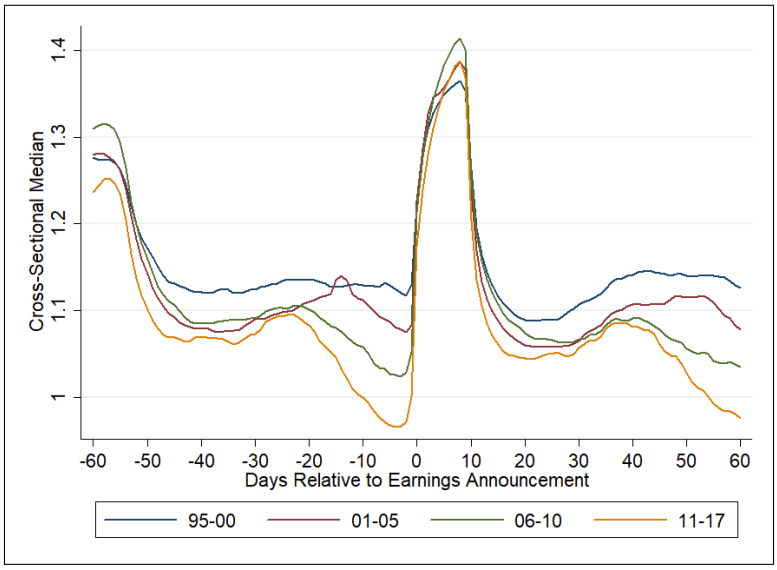

Figure 12. Decline of Pre-Earnings Volume, Expanded Window. Plot of 10-day moving average of abnormal volume. Abnormal volume is volume relative to the historical average over the past year. Average historical volume is fixed at the beginning of each 10-day moving-average window to avoid mechanically amplifying drops in volume.

The figure also motivates my choice of a 22 trading-day window for the drop in pre-earnings volume: This is where there are differences across years. This is a case of looking where the effect is, but nothing in this figure suggests that this trend is driven by changes in passive ownership.

Another explanation for decreased pre-earnings volume is that informed trading before earnings announcements has moved to dark pools. This could occur because on lit exchanges, informed traders are getting front-run by algorithm traders. To test this, I obtained data on dark pool volume from FINRA. There does not appear to be an increase in dark pool volume in the weeks before earnings announcements.

Another possibility is that the increase in algorithmic trader (AT) activity over time is driving the decrease in pre-earnings volume, as AT's may trade less before earnings announcements. In years where I can construct the AT activity measures of Weller (2017), and add them to the right-hand-side of Regression 2, there is still a statistically significant decline in volume before earnings announcements.

Appendix E. Pre-Earnings Drift

Table XI contains several alternative definitions of the pre-earnings drift. Only $DM_{i,t}$ is consistent with my intuition for all permutations of pre-earnings and earnings-day returns, and the pre-earnings information content of prices.

Case	$r_{i,(t-30,t-1)}$	$r_{i,t}$	$\frac{r_{i,t}}{r_{i,(t-30,t-1)}}$	$\frac{1+r_{i,t}}{1+r_{i,(t-30,t-1)}}$	$DM_{i,t}$	Intuition
1	-1.00%	-2.00%	2.00	1.01	0.33	Some Info.
2	-1.00%	2.00%	(2.00)	0.97	(1.00)	Low Info.
3	-1.00%	-0.50%	0.50	0.99	0.67	High Info.
4	1.00%	-2.00%	(2.00)	1.03	(1.00)	Low Info.
5	1.00%	2.00%	2.00	0.99	0.33	Some Info.
6	1.00%	0.50%	0.50	1.00	0.67	High Info.

Table XI Alternative of Pre-Earnings Drift Measures. This table presents 6 hypothetical scenarios. Columns 2 and 3 contain the pre-earnings and earnings-day returns. Columns 4 and 5 calculate the drift magnitude under alternative definitions in net and gross returns, while column 6 has my definition, $DM_{i,t}$.

Appendix F. Reduced-Form Placebo Regressions

This section contains placebo tests for the reduced-form regressions. I select dates between the actual earnings days to represent placebo earnings dates. For example, if a firm released earnings on 12/31/2017, I would select the trading day closest to 11/15/2017 as the placebo earnings date. In all cases, the results for placebo earnings days are insignificant.

Appendix G. Relationship to Competing Hypotheses

Appendix A. Rise of AT Activity

Tables XV, XVI and XVII contains alternative versions of the reduced-form regressions, which include controls for algorithmic trading (AT) activity.

Only the results for the pre-earnings drift and the earnings day share of volatility are significant in the matched subsample. For the specifications that are significant in the subset of my original sample that can be matched to the MIDAS data, adding the AT activity controls does reduce the coefficient on passive ownership/change in passive ownership, but the sign and statistical significance is unchanged. This implies that increased AT activity may partially explain the observed decrease in market efficiency, but passive ownership is still an important factor.

Appendix B. Regulation Fair Disclosure

Tables XVIII, XIX, XX contain alternative versions of the reduced-form regressions, restricting the sample to data after 2000. The results are qualitatively similar, which alleviates concerns of the results being driven by time trends resulting from Reg FD, which was passed in August 2000.

		1-year		3-year		5-year	
		Placebo	Baseline	Placebo	Baseline	Placebo	Baseline
Increase in Passive Ownership	1-year	-0.628	-1.263*				
		(0.691)	(0.651)				
	3-year			-0.67	-0.936***		
				(0.434)	(0.309)		
	5-year					-0.502	-1.041**
						(0.382)	(0.421)
Observations		239,402	272,609	157,530	191,654	109,632	139,537
R-Squared		0.016	0.009	0.024	0.007	0.047	0.025
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	No	No	No	No	No

Table XII Placebo Test: Pre-Earnings Volume. Estimates of β from:

$$\Delta_{(t,t-n)}CAV_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$ is average pre-earnings cumulative abnormal volume *per day* over the 22 days leading up to the earnings announcement. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Change in passive ownership is expressed as a decimal, so $0.01 = 1\%$ increase. Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry and year. The "Baseline" results are estimates from Table I, while the "Placebo" results are the coefficient estimates when selecting dates between the actual earnings days as the placebo earnings dates.

	Pre-Earnings Drift	
	Placebo	Baseline
Passive Ownership	-0.0133	-0.329**
	(0.071)	(0.157)
Observations	394,397	413,328
R-Squared	0.000	0.003
Firm Controls	Yes	Yes
SUE	Yes	Yes
Industry/Year FE	Yes	Yes
Firm FE	No	No

Table XIII Placebo Test: Pre-Earnings Drift. Table with estimates of β from:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=2}^{10} \phi_j \mathbf{1}_{SUEdecile=j} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, which is the ratio of the cumulative returns in the 30 days leading up to the earnings day, relative cumulative return in the 30 days up to and including the earnings day. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of shares outstanding held by passive funds. SUE deciles are formed each quarter. $DM_{i,t}$ is Winsorized at the 1% and 99% levels. Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry and year. All standard errors are clustered at the firm/year level.
The "Baseline" results are estimates from Table II, while the "Placebo" results are the coefficient estimates when selecting dates between the actual earnings days as the placebo earnings dates.

		1-year		3-year		5-year	
		Placebo	Baseline	Placebo	Baseline	Placebo	Baseline
Increase in Passive Ownership	1-year	0.00069	0.0983**				
		(0.010)	(0.039)				
	3-year			0.00329	0.103***		
				(0.005)	(0.034)		
	5-year					0.00818	0.171***
						(0.005)	(0.042)
Observations		65,951	79,253	50,151	60,578	40,215	48,484
R-Squared		0.002	0.015	0.002	0.052	0.003	0.077
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	No	No	No	No	No

Table XIV Placebo Test: Earnings Day Share of Volatility. Table with estimates of β from:

$$\Delta_{(t,t-n)}QVS_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + \gamma X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, which is the ratio of the squared returns on the 4 quarterly earnings announcement days, relative to the squared returns on all days in year t. QVS takes values in [0,1]. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry and year. All standard errors are clustered at the firm/year level.
The "Baseline" results are estimates from Table III, while the "Placebo" results are the coefficient estimates when selecting dates between the actual earnings days as the placebo earnings dates.

		1-year		3-year	
		Baseline	+AT Controls	Baseline	+AT Controls
Increase in Passive Ownership	1-year	-0.488	0.438		
		(0.961)	(0.913)		
	3-year			-0.496***	-0.253***
				(0.078)	(0.057)
Observations		38,081	38,048	15,681	15,665
R-Squared		0.006	0.054	0.044	0.125
Firm Controls		Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes
Firm FE		No	No	No	No

Table XV AT Activity: Pre-Earnings Volume. Estimates of β from:

$$\Delta_{(t,t-n)}CAV_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + \gamma X_{i,t-n} + \phi\Delta_{(t,t-n)}ATActivity_{i,t} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$ is average pre-earnings cumulative abnormal volume *per day* over the 22 days leading up to the earnings announcement. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Change in passive ownership is expressed as a decimal, so 0.01 = 1% increase. $\Delta_{(t,t-n)}ATActivity_{i,t}$ is a vector of year-over-year changes in the 4 AT activity measures from Weller (2017). Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry and year. Only includes data from 2012-2016.

	Pre-Earnings Drift	
	Baseline	+AT Controls
Passive Ownership	-1.281***	-1.074***
	(0.379)	(0.358)
Observations	54,119	54,079
R-Squared	0.004	0.004
Firm Controls	Yes	Yes
SUE	Yes	Yes
Industry/Year FE	Yes	Yes
Firm FE	No	No

Table XVI AT Activity: Pre-Earnings Drift. Table with estimates of β from:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=2}^{10} \phi_j \mathbf{1}_{SUEdecile=j} + \gamma X_{i,t-1} + \psi ATActivity_{i,t} + \text{Fixed Effects} + e_{i,t}$$

$DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, which is the ratio of the cumulative returns in the 30 days leading up to the earnings day, relative cumulative return in the 30 days up to and including the earnings day. Passive ownership is expressed as a decimal, so 0.01 = 1% of shares outstanding held by passive funds. SUE deciles are formed each quarter. $DM_{i,t}$ is Winsorized at the 1% and 99% levels. $ATActivity_{i,t}$ is a vector containing the 4 AT activity measures from Weller (2017). Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry and year. All standard errors are clustered at the firm/year level. Only includes data from 2012-2016.

		1-year		3-year	
		Baseline	+AT Controls	Baseline	+AT Controls
Increase in Passive Ownership	1-year	0.188**	0.167*		
		(0.089)	(0.092)		
	3-year			0.103***	0.0502*
				(0.034)	(0.026)
Observations		9,910	9,904	4,023	4,020
R-Squared		0.009	0.010	0.035	0.041
Firm Controls		Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes
Firm FE		No	No	No	No

Table XVII AT Activity: Earnings Day Share of Volatility. Table with estimates of β from:

$$\Delta_{(t,t-n)} QVS_{i,t} = \alpha + \beta \Delta_{(t,t-n)} Passive_{i,t} + \gamma X_{i,t-n} + \phi \Delta_{(t,t-n)} ATActivity_{i,t} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, which is the ratio of the squared returns on the 4 quarterly earnings announcement days, relative to the squared returns on all days in year t. QVS takes values in [0,1]. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. $\Delta_{(t,t-n)} ATActivity_{i,t}$ is a vector of year-over-year changes in the 4 AT activity measures from Weller (2017). Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry and year. All standard errors are clustered at the firm/year level. Only includes data from 2012-2016.

		1-year		3-year		5-year	
		(1)	(2)	(1)	(2)	(1)	(2)
Increase in Passive Ownership	1-year	-1.102*	-1.033***				
		(0.647)	(0.231)				
	3-year			-0.862***	-0.537***		
				(0.314)	(0.174)		
	5-year					-1.007**	-0.596***
						(0.434)	(0.188)
Observations		187,778	187,778	145,900	145,900	115,600	115,600
R-Squared		0.004	0.005	0.007	0.008	0.027	0.028
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

Table XVIII Post-2000: Pre-Earnings Volume. Estimates of β from:

$$\Delta_{(t,t-n)}CAV_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$ is average pre-earnings cumulative abnormal volume *per day* over the 22 days leading up to the earnings announcement. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Change in passive ownership is expressed as a decimal, so $0.01 = 1\%$ increase. Controls, $X_{i,t-n}$, include lagged passive ownership, market capitalization, idiosyncratic volatility and total institutional ownership. I also condition on the growth in market capitalization from $t-n$ to t. Fixed effects include 2-digit SIC industry and year. Only includes data from 2001-2016.

	Pre-Earnings Drift	
	(1)	(2)
Passive Ownership	-0.260**	-0.321***
	(0.102)	(0.096)
Observations	232,301	232,301
R-Squared	0.011	0.004
Firm Controls	Yes	Yes
SUE	Yes	Yes
Industry/Year FE	Yes	Yes
Firm FE	No	Yes

Table XIX Post-2000: Pre-Earnings Drift. Table with estimates of β from:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=2}^{10} \phi_j \mathbf{1}_{SUEdecile=j} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, which is the ratio of the cumulative returns in the 30 days leading up to the earnings day, relative cumulative return in the 30 days up to and including the earnings day. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of shares outstanding held by passive funds. SUE deciles are formed each quarter. $DM_{i,t}$ is Winsorized at the 1% and 99% levels. Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry and year. All standard errors are clustered at the firm/year level. Only includes data from 2001-2016.

		1-year		3-year		5-year	
		(1)	(2)	(1)	(2)	(1)	(2)
Increase in Passive Ownership	1-year	0.107***	0.0583				
		(0.041)	(0.038)				
	3-year			0.113***	0.0414		
				(0.037)	(0.032)		
	5-year					0.188***	0.0815**
						(0.044)	(0.033)
Observations		49,113	49,113	40,556	40,556	34,639	34,639
R-Squared		0.016	0.016	0.051	0.049	0.070	0.064
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Industry/Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

Table XX Post-2000: Earnings Day Share of Volatility. Table with estimates of β from:

$$\Delta_{(t,t-n)}QVS_{i,t} = \alpha + \beta\Delta_{(t,t-n)}Passive_{i,t} + \gamma X_{i,t-n} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, which is the ratio of the squared returns on the 4 quarterly earnings announcement days, relative to the squared returns on all days in year t. QVS takes values in [0,1]. $\Delta_{(t,t-n)}$ is the change from calendar year $t-n$ to calendar year t. Controls in $X_{i,t-1}$ include lagged institutional ownership, lagged market capitalization and lagged market-adjusted volatility over the past year. Fixed effects include 2-digit SIC industry and year. All standard errors are clustered at the firm/year level. Only includes data from 2001-2016.

Appendix H. Alternative Identified Evidence

Appendix A. S&P 500 Index Deletions

In section V, I use S&P 500 index additions to identify plausibly exogenous increases in passive ownership. A natural extension is to run a similar difference-in-differences regression, but use the decrease in passive ownership associated with index deletion as the treatment. In this DID setup, the parallel trends assumption is unlikely valid, because index deletion is always about firm fundamentals.

The next challenge is identifying the control group, which should consist of firms with a similar likelihood of being dropped from the index as the treated firms. Three major reasons for S&P 500 index deletion are small market capitalization, poor performance and lack of liquidity. To facilitate a direct comparison with the index addition results, I sort on industry, size and growth rate to identify control firms, even though removing the industry filter and replacing it with a measure of liquidity would probably yield a more appropriate control group.

In the index deletion setup, the treatment group is all firms dropped from the S&P 500 index. The control group is all firms in the same 2-digit SIC industry, in the same size and growth rate quintiles that were initially in the S&P 500 index, and remained there over the next two years.[24]

[24] All the results index deletion results are similar if the control group only includes firms were initially not in the S&P 500 index, and remained out of the index over the next two years. Results are also similar when choosing the treatment period to be the year immediately after index deletion, instead of skipping a year.

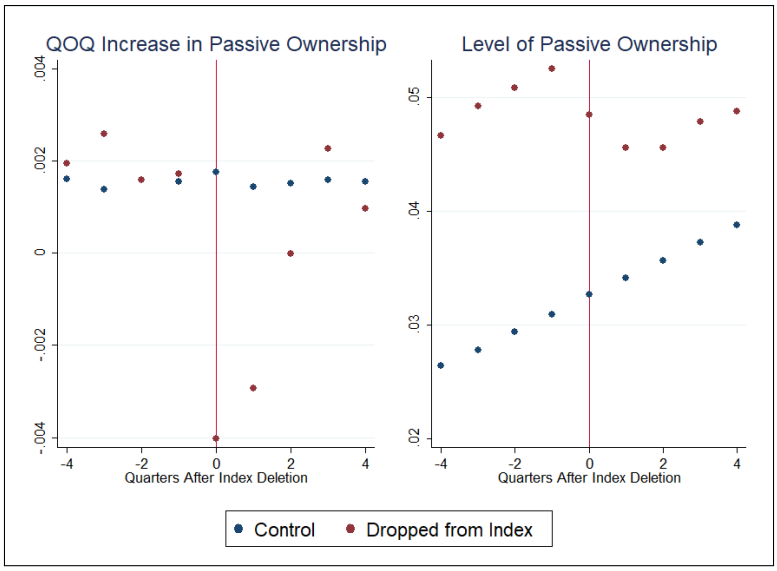

Figure 13. S&P 500 Index Deletions: Testing Parallel Trends. Average level and increase in passive ownership for control firms and firms dropped from the S&P 500. Control firms are all firms in the same 2-digit SIC industry, in the same size and growth rate quintiles that were initially in the S&P 500 index, and remained there over the next two years.

Figure 13 shows the changes in passive ownership around the index deletion date. There is a drop in passive ownership in the quarter of deletion, and the quarter after deletion. Unlike the increase in passive ownership after index addition, however, the decrease after index deletion is only temporary, as can be seen in the levels plot. One explanation for this is that stocks on the margin are still relatively large, and were affected by the ETF/passive management boom which increased passive ownership for all stocks. The weak and temporary treatment effect suggests that the DID results will be insignificant.

The regressions testing the effect of index deletions on pre-earnings volume, drift and volatility are in Table XXI. I omit the implied elasticity calculation because the first stage effect is near zero. The pre-earnings drift and earnings day share of volatility are now insignificant, while the volume result is significant and negative.

Because the treated firms are experiencing a *decrease* in passive ownership, this volume result appears to be contrary to the results in the main body of the paper. Some possible explanations for this are that S&P 500 funds have (1) a different propensity to lend out shares (2) a higher share of volume coming from algorithmic traders (because of ETF arbitrage, for example).

Appendix B. Moving from the Russell 2000 to the Russell 1000

Similar to the S&P 500 index deletion, firms experience a decrease in passive ownership after they are moved from the Russell 2000 to the Russell 1000. This is because they go from being the largest firm in a value-weighted index of small firms, to the smallest firm in a value-weighted index of large firms. Unlike the S&P 500 deletions, however, this DID setup still satisfies the exclusion

	Pre-Earnings Volume Differences	Pre-Earnings Drift Levels	Earnings Day Share of QV Differences
Treated	-0.0699*		-0.0197
	(0.026)		(0.017)
Interaction		-0.0742	
		(0.100)	
R-squared	0.076	0.028	0.23
Year/Month FE	YES	YES	YES
Industry FE	YES	YES	YES
Size Quintile FE	YES	YES	YES
Growth Quintile FE	YES	YES	YES
Treated Firms	245	245	245
Control Firms	1,239	1,239	1,239

Table XXI Effects of S&P 500 Index Deletion. For Pre-Earnings Volume and Earnings Day Share of QV:

$$\Delta_{(t-1,t+1)} Outcome_{i,t} = \alpha + \gamma Dropped_{i,t} + \text{Fixed Effects} + e_{i,t}$$

For Pre-Earnings Drift:

$$Outcome_{i,t} = \alpha + \beta Dropped_{i,t} + \tau After_{i,t} + \gamma Added_{i,t} \times After_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Dropped firms are those which were removed to the S&P 500 index. Control firms are in the same industry, same size and same growth rate quintile as the control firms. Fixed effects include industry, size quintile, growth rate quintile, and year/month of index deletion.

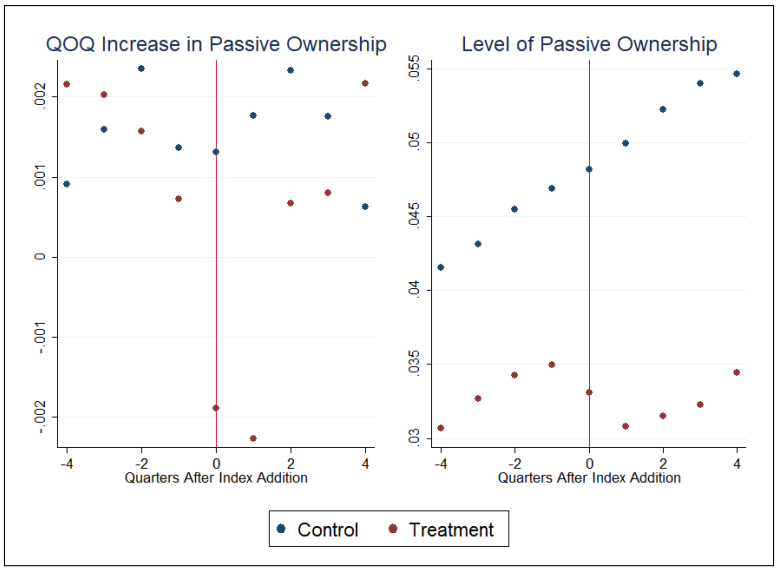

Figure 14. Russell 2000/1000 Reconstitution: Checking Parallel Trends. Average level and change in passive ownership for control firms and firms moved from the Russell 2000 to the Russell 1000. Control firms are all firms in the Russell 3000 ranked 900 to 1100 that did not move from the 1000 to the 2000 or from the 2000 to the 1000.

restriction, as moving from firm 1001 to 999 may have nothing to do with firm fundamentals.

Similar to the setup in Section V, I choose the control firms to be all Russell 3000 firms, with June ranks between 900 and 1100 that did not switch from the 1000 to the 2000, or from the 2000 to the 1000. Figure 7 shows the problem with this setup: the treatment is small and temporary. The common trend between moving from the Russell 2000 to the 1000, and the S&P 500 index deletion suggests that the general upward trend in passive ownership for almost all stocks drowns out the temporary change in passive ownership associated with various index rebalancing.

In Table XXII, I replicate the Russell experiment results of Table VIII. As with the S&P 500 deletions, the pre-earnings drift and earnings day share of volatility are now insignificant, while the volume result is significant. Again, this volume result is contrary to the results in the main body of the paper – suggesting there is not something special about deletion from the S&P 500 index and the associated drop in pre-earnings volume.

Appendix I. Systematic Information Announcement Days

I obtain FOMC announcement dates from Gorodnichenko and Weber (2016). To account for the anticipated nature of earnings announcements, I restrict the sample to scheduled FOMC meetings.

Figure 15 shows the trends in volume before FOMC announcement dates. The periodic oscillation is due to day of the week effects. Figure 16 shows the pre-FOMC announcement drift, which, if anything, has been trending up. Figure 17 shows a slight trend toward increased volatility on FOMC announcement dates, but this may be due to the increased importance of FOMC meetings during the financial crisis.

	Pre-Earnings Volume Differences	Pre-Earnings Drift Levels	ED Share of QV Differences
Treated	-0.123***		0.00498
	(0.021)		(0.006)
Interaction		-0.0531	
		(0.040)	
R-squared	0.016	0.006	0.040
Year/Month FE	YES	YES	YES
Treated Firms	974	974	974
Control Firms	1,284	1,284	1,284

Table XXII Effects of Russell 1000/2000 Index Reconstitution. For Pre-Earnings Volume and Earnings Day Share of QV:

$$\Delta_{(t-1,t+1)} Outcome_{i,t} = \alpha + \gamma Moved_{i,t} + \text{Fixed Effects} + e_{i,t}$$

For Pre-Earnings Drift:

$$Outcome_{i,t} = \alpha + \beta Moved_{i,t} + \tau After_{i,t} + \gamma Moved_{i,t} \times After_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Added firms are those which were moved from the Russell 2000 to the Russell 1000. Control firms are those that were ranked 900-1100 by Russell, but did not switch from the 1000 to the 2000 or from the 2000 to the 1000. Fixed effects include year/month of index rebalancing.

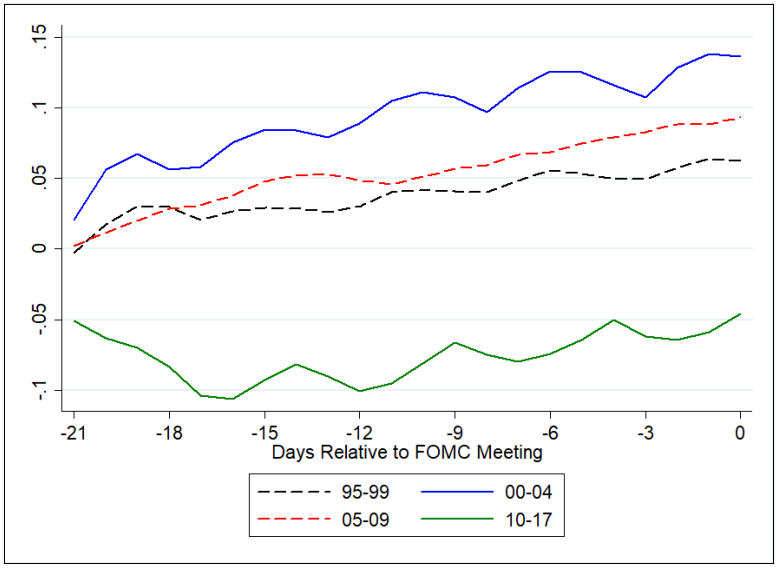

Figure 15. FOMC Meeting Dates: Pre-Earnings Volume. Plot of $\beta_{\{j==-\tau\}}$ estimated from the regression:

$$\overline{CAV_{i,j,t}} = \alpha + \sum_{\tau=-21}^{0} \beta_\tau \mathbf{1}_{\{j==-\tau\}} + \text{Fixed Effects} + e_{i,j,t}$$

Where t denotes a scheduled FOMC announcement date.

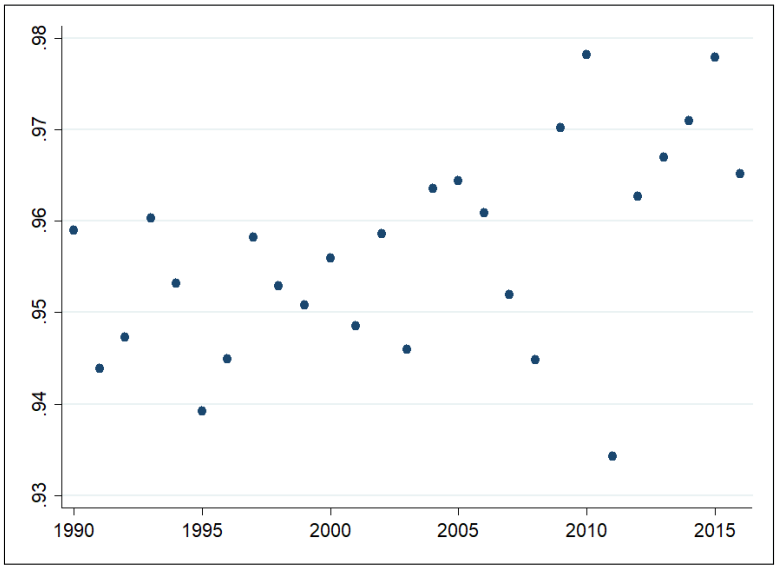

Figure 16. FOMC Meeting Dates: Pre-Earnings Drift. This figure plots the cross-sectional average of the drift magnitude measure, $DM_{i,t} = \frac{r_{i,(t-30,t-1)}}{r_{i,(t-30,t)}}$, by year where t denotes a scheduled FOMC meeting date.

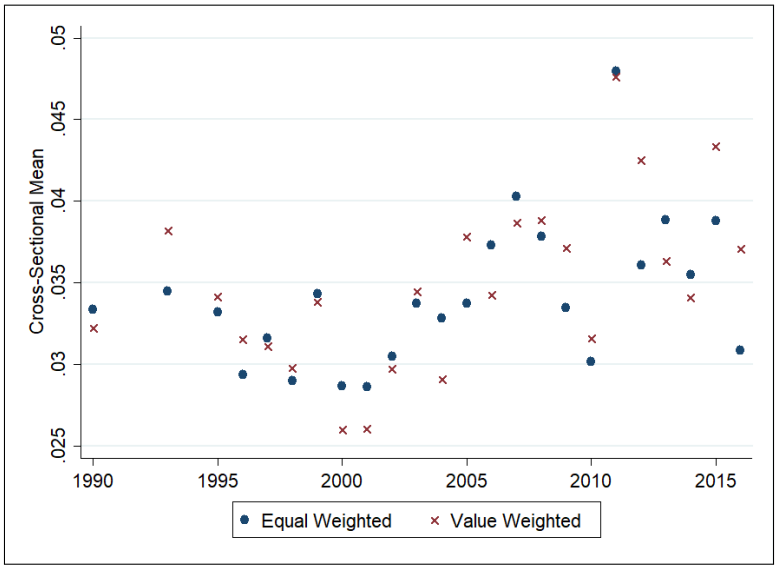

Figure 17. FOMC Meeting Dates: Earnings Day Volatility. This figure plots the share of market-adjusted quadratic variation occurring on placebo earnings days. For firm i in year t the quadratic variation share (QVS) is defined as: $QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, where r denotes a market-adjusted daily return. The numerator is the sum of squared returns on the 8 scheduled FOMC dates , while the denominator is the sum of squared returns on all trading days in year t.

Appendix J. Model Details

In this Appendix, I discuss the details of the Grossman and Stiglitz (1980)-style model from Section II, and show various comparative statics. This section borrows heavily from Dimitris Papainkolaou's teaching notes, *"Models with Asymmetric Information"*.

Appendix A. Model Setup

Consider a model with two periods, 0 and 1. There is a single risky asset that pays a liquidating dividend of d in period 1, with total supply S. Assume d is normally distributed with mean $\overline{d}$ and variance σ_d^2. There are three types of agents: N_I of them are informed, and get a private signal $s = d + \epsilon$ in period 0, where ϵ is normally distributed with mean zero and variance σ_ϵ^2. All of the informed agents receive the same signal s. N_U are uninformed, and only learn about the liquidating dividend through the price. Both the informed and uninformed have exponential utility over period 1 consumption, with coefficients of absolute risk aversion α_I and α_U.

The third type of agents are noise traders who buy and sell u shares with no regard for the price, where u is normally distributed with mean zero and variance σ_u^2. This implies that the effective supply of the asset is $S + u$. The presence of noise traders prevents the uninformed agents from perfectly learning the informed agents' signal from the price in a rational expectations equilibrium.

Denote the price of the asset in period zero as p, the risk-free rate between period 0 and 1 as r and initial wealth as W_0. If an informed or uninformed agent buys x shares of the risky asset at time 0, her period one wealth will be:

$$W_1 = (W_0 - xp)(1 + r) + xd$$

And expected utility will be:

$$-E\left[e^{-\alpha_i(W_0 - xp)(1+r)+xd}\right] \tag{J1}$$

Because d is normal, maximizing J1 is equivalent to maximizing:

$$(W_0 - xp)(1 + r) + xE[d|I_i] - \frac{1}{2}\alpha_i \sigma^2_{d|I_i} x^2 \tag{J2}$$

where I_i is the information set of agent i. For the informed agent, this will be her private signal and the price. For the uninformed agent, I_i will be the price.

The rational expectations equilibrium is a price function, $p(s, u)$ and a vector of demands, $\left(x^1(p), \ldots, x^{N_I}(p), x^{N_I+1}(p), \ldots, x^{N_I+N_U}(p)\right)$ which satisfy optimality and market clearing.

From J2, we can solve for the demand of the informed given the price, $x_I(p)$ and the demand of the uninformed given the price $x_U(p)$. Then, we substitute the demands into the market clearing condition:

$$N_I x_I(p) + N_U x_U(p) = S + u \tag{J3}$$

and finally solve for the equilibrium price, as a function of s, u and model parameters. If $\alpha_U =$

$\alpha_I = \alpha$, the price function takes the form:

$$price = A + B\left[(s - \overline{d}) - Cu\right] \tag{J4}$$

where A, B and C are constants that depend on model parameters:

$$A = \left(\frac{\overline{d}}{1+r} - \frac{S\alpha\sigma^2_{d|s}}{(1+r)\left(N_I + N_U \frac{\sigma^2_{d|s}}{\sigma^2_{d|p}}\right)} \right)$$

$$B = \frac{N_I \beta_s}{(1+r)\left(N_I + N_U \frac{\sigma^2_{d|s}}{\sigma^2_{d|p}}\right) - N_U \frac{\sigma^2_{d|s}}{\sigma^2_{d|p}} \beta_p} \tag{J5}$$

$$C = \frac{\alpha\sigma^2_\epsilon}{N_I}$$

where $\beta_p = \frac{\sigma^2}{B(\sigma^2 + \sigma^2_\epsilon + C^2\sigma^2_u)}$. To solve for the equilibrium price, start by solving for A and C, then jointly solve for B and β_p.

Appendix B. Volatility of Earnings Day Returns

Given the price function, we can compute the conditional mean of d given the price p:

$$E[d|p] = \overline{d} + \frac{\sigma^2}{B(\sigma^2 + \sigma^2_\epsilon + C^2\sigma^2_u)}(p - A) \tag{J6}$$

Then, we can compute the conditional variance of d using the law of total variance: $V(d) = E[V(d|p)] + V(E[d|p])$. Rearranging yields:

$$\sigma^2_{d|p} = V(d) - V(E[d|p]) = \sigma^2 - \frac{\sigma^4}{\sigma^2 + \sigma^2_\epsilon + C^2\sigma^2_u} \tag{J7}$$

Informativeness is the inverse of the conditional variance of d:

$$informativeness = \frac{1}{\sigma^2_{d|p}} = \frac{\sigma^2_d + \sigma^2_\epsilon + \left(\frac{\alpha\sigma^2_\epsilon}{N_I}\right)^2 \sigma^2_u}{\sigma^2_d \left(\sigma^2_\epsilon + \left(\frac{\alpha\sigma^2_\epsilon}{N_I}\right)^2 \sigma^2_u\right)} \tag{J8}$$

The derivative of informativeness with respect to N_I is:

$$\frac{2\alpha^2\sigma^4_\epsilon\sigma^2_u N_I}{\left(\sigma^2_\epsilon N_I^2 + \alpha^2\sigma^4_\epsilon\sigma^2_u\right)^2} \tag{J9}$$

Informativeness always increasing in N_I.

At $t = 1$, uncertainty is resolved, which we can equate to an earnings announcement. In Section III, I show that the volatility on earnings announcement dates, relative to non-announcement dates, is increasing in the share of passive ownership. We can view the expected return variance as an analogue to average earnings day volatility.

Taking the interpretation that passive owners are uninformed, define the uninformed share as $\frac{N_U}{N_I+N_U}$. Define the asset's return as the percentage difference between the $t = 0$ price p and the terminal dividend d. This is a function of d, ϵ (the noise in the signal), u (the volume of noise trading), N_I (the number of informed traders) and N_U (the number of uninformed traders): $r(d, \epsilon, u, N_I, N_U) = \frac{d-p}{p}$. Because the distributions of d, ϵ and u are known, we can compute the expected return variance as:

$$Var(r|N_I, N_U) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r(d, \epsilon, u, N_I, N_U)^2 dF(\epsilon)dF(d)dF(u) - \left(\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} r(d, \epsilon, u, N_I, N_U)^2 dF(\epsilon)dF(d)dF(u)\right)^2 \tag{J10}$$

I evaluate this integral numerically by drawing one million observations from the distributions of ϵ, d and u.

Figure 18 shows that as the share of uninformed investors increases, the expected return variance increases, and that this effect is non-linear. When the uninformed share is low, the marginal effect of increases in uninformed is small, but when the uninformed share is high, the marginal effect of increases in uninformed is amplified.

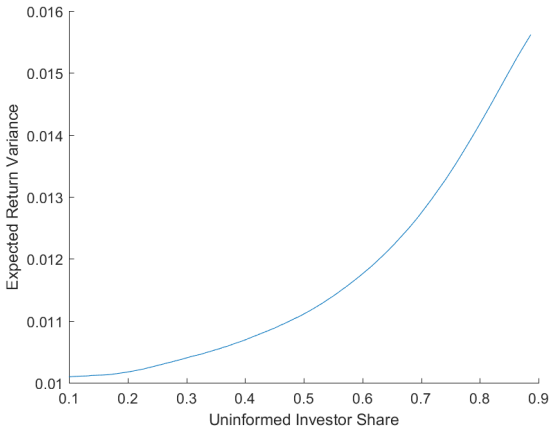

Figure 18. Passive Share and Expected Return Variance. This figure plots the relationship between the uninformed share, $\frac{N_U}{N_I+N_U}$, and the expected variance of returns. Parameters: $r = 0.05$, $\overline{d} = 2$, $\sigma^2 = 0.2$, $\sigma_\epsilon^2 = 0.3$, $\sigma_u^2 = 0.4$, $S = 1$, $\alpha = 3$

The derivative of informativeness with respect to σ_u^2 is:

$$-\frac{\alpha^2\sigma_\epsilon^4 N_I^2}{\left(\sigma_\epsilon^2 N_I^2 + \alpha^2\sigma_\epsilon^4\sigma_u^2\right)^2} \tag{J11}$$

Informativeness always decreasing in σ_u^2. Taking the interpretation that increases in passive ownership translate to increases in noise trading, we can fix the passive share, and look at the effect of increasing the variance of noise trading on the variance of returns, again using Equation J10.

Figure 19 shows that as the variance of noise trading increases, the expected return variance increases. Like the response of price informativeness to increasing the share of uninformed investors, this effect is non-linear.

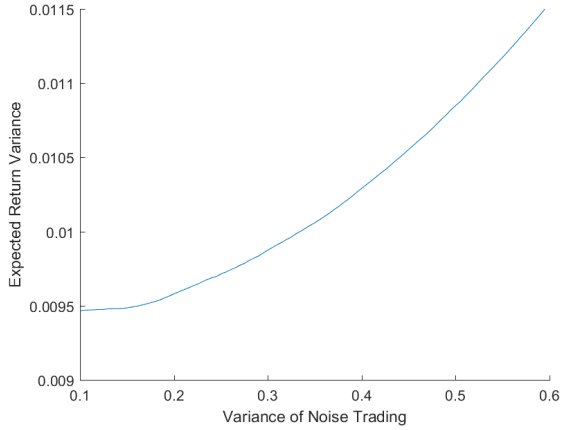

Figure 19. Variance of Noise Trading and Expected Return Variance. This figure plots the relationship between the variance of noise trading, σ_u^2, and the expected variance of returns. Parameters: $r = 0.05$, $\overline{d} = 2$, $\sigma^2 = 0.2$, $\sigma_\epsilon^2 = 0.3$, $\frac{N_U}{N_I+N_U} = 0.1$, $S = 1$, $\alpha = 3$

All of the analysis so far assumes that the mean noisy demand, $E[u]$, is zero. Suppose that instead of increasing the variance of noise trading, increases in passive ownership translate to a higher mean noisy demand. Figure 20 shows that as the mean of noisy demand increases, the expected return variance increases and the relationship is linear.

In all the previous analyses, I assumed that the risk aversion for informed and uninformed agents was the same. We can re-solve the model, except allow the informed investor's risk aversion, α_I to be different from the uninformed investor's risk aversion, α_U.

The price still takes the form:

$$price = \tilde{A} + \tilde{B}\left[(s - \overline{d}) - \tilde{C}u\right] \tag{J12}$$

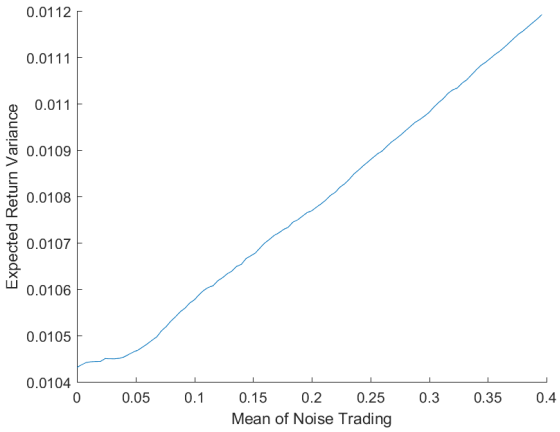

Figure 20. Mean of Noise Trading and Expected Return Variance. This figure plots the relationship between the mean of noise trading, and the expected variance of returns. Parameters: $r = 0.05$, $\overline{d} = 2$, $\sigma^2 = 0.2$, $\sigma_\epsilon^2 = 0.3$, $\frac{N_U}{N_I+N_U} = 0.1$, $S = 1$, $\alpha = 3$, $\sigma_u^2 = 0.4$.

but $\tilde{A}$, $\tilde{B}$ and $\tilde{C}$ now reflect the different risk aversions:

$$\tilde{A} = \frac{\overline{d}}{1+r} - \frac{S}{(1+r)\left(\frac{N_I}{\alpha_I \sigma^2_{d|s}} + \frac{N_U}{\alpha_U \sigma^2_{d|p}}\right)}$$

$$\tilde{B} = \frac{\frac{N_I \beta_s}{\alpha_I \sigma^2_{d|s}}}{N_I \frac{1+r}{\alpha_I \sigma^2_{d|s}} - N_U \frac{\tilde{\beta}_p - (1+r)}{\alpha_U \sigma^2_{d|p}}} \tag{J13}$$

$$\tilde{C} = \frac{\alpha_I \sigma_\epsilon^2}{N_I}$$

where $\tilde{\beta}_p = \frac{\sigma^2}{\tilde{B}(\sigma^2+\sigma_\epsilon^2+\tilde{C}^2\sigma_u^2)}$. To solve for the equilibrium price, first solve for $\tilde{A}$ and $\tilde{C}$, then jointly solve for $\tilde{B}$ and $\tilde{\beta}_p$.

Figure 21 shows the effect of decreasing the risk aversion of the informed agents, α_I, while keeping the risk aversion of the uninformed agents, α_U, constant. Decreasing the informed agent's risk aversion decreases the expected return variance.

Appendix K. Investment Q Relationship

Q-theory proposes a positive relationship between marginal Q and investment. If passive ownership has made prices less informative, the market value of the firm has become a less accurate measure of the true value of the firm. This implies that investment should become less sensitive to Q for firms with high passive ownership.

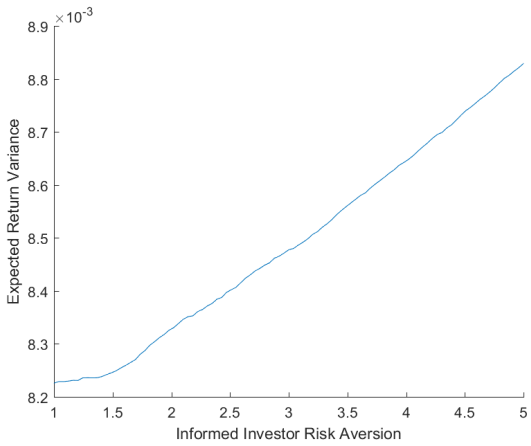

Figure 21. Informed Investor Risk Aversion and Expected Return Variance. This figure plots the relationship between the risk aversion of informed traders, and the expected variance of returns. Parameters: $r = 0.05$, $\overline{d} = 2$, $\sigma^2 = 0.2$, $\sigma_\epsilon^2 = 0.3$, $\frac{N_U}{N_I + N_U} = 0.1$, $S = 1$, $\alpha_U = 3$, $\sigma_u^2 = 0.4$.

I test this in reduced form, and using plausibly exogenous increase in passive ownership that arises from being added to the S&P 500 index. Because marginal Q is hard to measure, I will work with average Q, the market-to-book ratio. The reduced-form specification is based on Eberly, Rebelo, Vincent, et al. (2008):

$$\left(\frac{I}{K}\right)_{i,t} = \alpha + \beta_1 Log(Q)_{i,t} + \beta_2 Passive_{i,t} + \beta_3 Passive_{i,t} \times Log(Q)_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{K1}$$

where K is the replacement value of capital, calculated using the method in Salinger and Summers (1983). Q= (Market Value of Equity + Book Value of Debt - Book Value of Inventories)/Capital. The S&P addition specification is identical to the difference-in-differences specification in Section V, except $Log(Q)$ is also included on the right-hand-side as a control. Fixed effects include industry, year and firm. The coefficient of interest is β_3, which should be negative if firms with high passive ownership have investment that is less sensitive to average Q.

Table XXIII contains the regression results. Consistent with decreased price informativeness, the investment of firms with a higher share of passive ownership is less sensitive to average Q.

	Investment/Capital		
Log(Q)	0.0371***	0.0394***	0.0418***
	(0.001)	(0.001)	(0.005)
Passive Share		0.1733***	
		(0.025)	
Passive X Log(Q)		-0.0861***	
		(0.009)	
Log(Q) x Post Add to S&P 500			-0.0140*
			(0.008)
Observations	76,227	76,227	2,574
R-squared	0.063	0.065	0.49
Year FE	YES	YES	See S&P 500
Industry FE	YES	YES	Addition Setup
Firm FE	YES	YES	

Table XXIII Passive Ownership and the Investment-Q Relation. Table with estimates of β_1, β_2, and β_3 from:

$$\left(\frac{I}{K}\right)_{i,t} = \alpha + \beta_1 Log(Q)_{i,t} + \beta_2 Passive_{i,t} + \beta_3 Passive_{i,t} \times Log(Q)_{i,t} + \text{Fixed Effects} + e_{i,t}$$

where K is the replacement value of capital, calculated using the method in Salinger and Summers (1983). Q= (Market Value of Equity + Book Value of Debt - Book Value of Inventories)/Capital. The S&P addition specification is identical to the difference-in-differences specification in Section V, except $Log(Q)$ is also included on the right-hand-side as a control. Fixed effects include industry, year and firm.

REFERENCES

Abdi, Farshid, and Angelo Ranaldo, 2017, A simple estimation of bid-ask spreads from daily close, high, and low prices, *The Review of Financial Studies* 30, 4437–4480.

Anson, Mark JP, Donald R Chambers, Keith H Black, Hossein Kazemi, CAIA Association, et al., 2012, *CAIA level I: an introduction to core topics in alternative investments*, volume 709 (John Wiley & Sons).

Appel, Ian R, Todd A Gormley, and Donald B Keim, 2016, Passive investors, not passive owners, *Journal of Financial Economics* 121, 111–141.

Azar, José, Martin C Schmalz, and Isabel Tecu, 2018, Anticompetitive effects of common ownership, *The Journal of Finance* 73, 1513–1565.

Buffa, Andrea M, Dimitri Vayanos, and Paul Woolley, 2014, Asset management contracts and equilibrium prices, Technical report, National Bureau of Economic Research.

Campbell, John Y, Martin Lettau, Burton G Malkiel, and Yexiao Xu, 2001, Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk, *The Journal of Finance* 56, 1–43.

Cremers, KJ Martijn, and Antti Petajisto, 2009, How active is your fund manager? a new measure that predicts performance, *The Review of Financial Studies* 22, 3329–3365.

Eberly, Janice C, Sergio Rebelo, Nicolas Vincent, et al., 2008, Investment and value: A neoclassical benchmark, Technical report, National Bureau of Economic Research.

Engelberg, JOSEPH, and Pengjie Gao, 2011, In search of attention, *The Journal of Finance* 66, 1461–1499.

Farboodi, Maryam, and Laura Veldkamp, 2017, Long run growth of financial technology, Technical report, National Bureau of Economic Research.

Glosten, Lawrence R, Suresh Nallareddy, and Yuan Zou, 2016, Etf trading and informational efficiency of underlying securities .

Gorodnichenko, Yuriy, and Michael Weber, 2016, Are sticky prices costly? evidence from the stock market, *American Economic Review* 106, 165–99.

Grossman, Sanford J, and Joseph E Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American economic review* 70, 393–408.

Israeli, Doron, Charles MC Lee, and Suhas A Sridharan, 2017, Is there a dark side to exchange traded funds? an information perspective, *Review of Accounting Studies* 22, 1048–1083.

Loughran, Tim, and Bill McDonald, 2017, The use of edgar filings by investors, *Journal of Behavioral Finance* 18, 231–248.

Mauboussin, Michael, Dan Callahan, and Darius Majd, 2017, Looking for easy games, Technical report, Credit Suisse.

McLean, R David, and Jeffrey Pontiff, 2016, Does academic research destroy stock return predictability?, *The Journal of Finance* 71, 5–32.

Morck, Randall, and Fan Yang, 2001, The mysterious growing value of s&p 500 membership, Technical report, National Bureau of Economic Research.

Novy-Marx, Robert, 2015, Fundamentally, momentum is fundamental momentum, Technical report, National Bureau of Economic Research.

Salinger, Michael, and Lawrence H Summers, 1983, Tax reform and corporate investment: A microeconometric simulation study, in *Behavioral simulation methods in tax policy analysis*, 247–288 (University of Chicago Press).

S&P, Dow Jones Indices, 2017, S&p u.s. indices methodology, Technical report.

Wang, Jiang, 1994, A model of competitive stock trading volume, *Journal of political Economy* 102, 127–168.

Wei, Wei, and Alex Young, 2017, Selection bias or treatment effect? a re-examination of russell 1000/2000 index reconstitution .

Weller, Brian M, 2017, Does algorithmic trading reduce information acquisition?, *The Review of Financial Studies* hhx137.