

# Passive Ownership and Price Informativeness

MARCO SAMMON\*

September 17, 2020

## ABSTRACT

This paper incorporates passive ownership into a model with endogenous learning. Passive ownership affects both how many investors decide to become informed, and what risks informed investors learn about. The model guides three new measures of price informativeness around days we know information is released: earnings announcements. Between 1990 and 2018, pre-earnings trading volume and pre-earnings drift declined, while volatility on earnings days increased. At the firm-level, there is a negative relationship between passive ownership and pre-earnings price informativeness. This result is robust to using only quasi-exogenous increases in passive ownership associated with S&P 500 additions and Russell 1000/2000 rebalancing.

JEL classification: G12, G14.

---

\*Kellogg School of Management. Email: [mcsammon@gmail.com](mailto:mcsammon@gmail.com). Website: [marcosammon.com](http://marcosammon.com). I would like to thank Scott Baker, Nick Bloom, Alex Chincó, Ian Dew-Becker, Ravi Jagannathan, Robert Korajczyk, Dimitris Papanikolaou, Martin Thyrsgaard and participants at the Kellogg Bag Lunch for helpful comments. This paper previously circulated as *Earnings Announcements and the Rise of Passive Ownership and ETFs, Learning, and Information in Stock Prices*

# 1 Introduction

The rise of passive ownership is one of the biggest changes in asset markets over the past 30 years. Passive funds grew from owning less than 1% of the US stock market in the early 1990s to owning nearly 15% in 2018. As passive ownership continues to grow, academics and practitioners want to understand: How does passive ownership affect stock price informativeness?

There is no consensus answer to this question in the theoretical or applied literature. Glosten et al. (2016) and Cong and Xu (2016) argue that passive ownership makes prices more informative through increased incorporation of systematic information. On the other hand, Ben-David et al. (2018) and Kacperczyk et al. (2018) show that passive ownership increases non-fundamental volatility and makes prices less informative<sup>1</sup>. A casual intuition in the popular press is that as more investors become passive, there are fewer investors left to do fundamental research and prices should become less informative. I document a new stylized fact consistent with this intuition: prices have become less informative before earnings announcements over the past 30 years.

Figure 1 shows the dynamics of cumulative abnormal returns (left panel) and abnormal trading volume (right panel) around earnings announcements. The sample includes firms in the top decile of standardized unexpected earnings (SUE) i.e., firms that had the best earnings news. In the early 1990s (solid blue line), prices trend up significantly before the good news is released, and there is no slow-down in trading. The return on the earnings day itself is small, relative to the run-up over the previous 30 days. The originators of these techniques (e.g. Ball and Brown (1968), Fama et al. (1969)) would likely argue that this is evidence of informed investors trading fundamental information into prices before it is formally announced.

[Figure 1 about here.]

Compare these patterns to what we see after 2010 (green line): The pre-earnings drift is smaller, and the move on earnings days is larger, relative to the pre-earnings drift. Further, investors are trading less in the weeks before the announcement and trading heavily after

---

<sup>1</sup>See also Israeli et al. (2017), Bhattacharya and O'Hara (2018), Malikov (2018), Garleanu and Pedersen (2018) and Chinco and Fos (2019).

the information is made public. From this comparison, it appears that prices were more informative before earnings announcements in the early 1990s, when passive ownership was negligible, than they are now, when passive ownership is large. Given that the most dramatic changes happen after the 2000's (dashed red line), it is unlikely that this trend was driven entirely by Regulation Fair Disclosure (passed in August 2000), and changes in the enforcement of insider trading laws (see Coffee Jr (2007)).

Two trends are not causal. A natural starting point to formalize the relationship between these two trends is Grossman and Stiglitz (1980): In this model, price informativeness decreases as the share of uninformed investors increases. The issue is that Grossman-Stiglitz does not necessarily apply to the rise of passive ownership. First, passive ownership is not necessarily uninformed. According to Daniel Gamba, former head of BlackRock's ETF business, only 30% of, "*ETF investors look at these as passive funds, [and] are just there long term.*"<sup>2</sup>

Passive ownership may also affect who acquires information, and what investors acquire information about. Many ETF holders are sophisticated institutional investors looking for targeted factor exposure. In reference to Global X's ETF offerings, former CEO Bruno del Ama said, "*Hedge funds tend to use our ETFs as a tactical play to get in and out of segments that are difficult for them to access directly. Greece is a good example. GREK has seen a lot [of] hedge fund trading.*"<sup>3</sup>

In this paper, I develop a model to address these issues. Passive ownership modeled as the fraction of a stock's shares outstanding held by an ETF. This ETF is traded by both informed and uninformed investors. The model also features two endogenous learning margins: (1) the extensive margin, which is the decision to pay a fixed cost and become informed or stay uninformed and (2) the intensive margin, which is the informed investors decision about how to allocate limited attention between systematic and stock-specific risks.

Increasing passive ownership in the model affects both the extensive and intensive learning margins. The signs of these effects, however, are ambiguous because of three competing channels. The first is the hedging channel: Passive ownership makes it easier for informed investors to take targeted bets on individual securities, because they can hedge out systematic risk with the ETF. This tends to increase the share of informed agents, and increases at-

---

<sup>2</sup>Daniel Gamba, Global Head of Active Equity Product Strategy, BlackRock, quoted in Balchunas (2016).

<sup>3</sup>Bruno del Ama, former CEO of Global X, quoted in Balchunas (2016).

tention on stock-specific risks. The second is the market-timing channel: Passive ownership allows investors to directly bet on the systematic risk-factor, which tends to increase attention on systematic risk. The third is the diversification channel: Passive ownership makes uninformed investors better off by giving them easy access to a well-diversified portfolio. This tends to increase the share of uninformed investors.

Which channels dominate depend on model parameters. The natural next step is to calibrate the model to match the data. The share of informed investors, and what investors learn about, however, is not easily observable. I focus on the models predictions for quantities easily measured in the data: returns, volume and volatility. I use these predictions to define three measures of price informativeness: (1) pre-earnings trading volume (2) the pre-earnings drift and (3) earnings-day volatility.

As the share of informed investors decreases, or attention on stock-specific risk decreases, all three price informativeness measures decrease as well. Like the intensive and extensive learning margins, however, the model has ambiguous predictions for the effect of passive ownership on price informativeness. I create empirical analogues of these three price informativeness measures, and show that they declined on average over the past 30 years.

In cross-sectional regressions, I find that increases in passive ownership are associated with a drop in trading volume before earnings announcement dates. Passive ownership is also correlated with decreased pre-earnings drift, and increased volatility on earnings days. All these regressions include year/quarter fixed-effects, so they are effectively comparing firms with different increases in passive ownership at the same point in time. This alleviates concerns that the results are driven by concurrent trends, or regime shifts in financial markets, such as changes in pre-earnings information leakages and the enforcement of insider trading laws. These reduced-form results imply that passive ownership decreases price informativeness.

The model can be calibrated to quantitatively match the observed rise in passive ownership, and qualitatively match the cross-sectional regression results. I also provide direct evidence of the effect of passive ownership on learning. Increases in passive ownership are correlated with decreased analyst coverage, less accurate analyst forecasts, more time between analyst updates and decreased download of SEC filings. This, however, raises the concern of reverse causality: perhaps passive ownership happened to increase the most in stocks that had the biggest decrease in price informativeness for other reasons.

To rule out reverse causality, I replicate my baseline regressions using only quasi-exogenous increases in passive ownership that arise from S&P 500 index additions and Russell 1000/2000 index rebalancing. All of the baseline results are qualitatively unchanged in these better-identified settings. These regressions include month-of-index-addition fixed effects, further ruling out the possibility that my results are driven by simultaneous trends or regime shifts.

**Paper Outline.** Section 2 sets up the model, and outlines the predicted effects of passive ownership on learning and price informativeness. Section 3 maps the model-based measures of price informativeness to the data. It also shows a decrease in average pre-earnings price informativeness between 1990 and 2018. Section 4 links the trends in passive ownership and price efficiency through cross-sectional regressions. Section 5 uses S&P 500 index additions and Russell 1000/2000 index rebalancing to identify increases in passive ownership which are plausibly uncorrelated with firm fundamentals. Price informativeness also decreases for firms with these quasi-exogenous increases in passive ownership.

## 2 Model of learning and passive ownership

In this section, I incorporate passive ownership into a model with endogenous learning. Investors face two learning decisions: (1) whether or not to pay a fixed cost to receive signals about asset payoffs and (2) how to allocate their limited attention, which determines the how precise these signals are for different assets' payoffs. The effect of increasing passive ownership on both learning decisions is ambiguous. Although learning is hard to measure empirically, the model has testable predictions for quantities directly observable in the data: trading volume, returns and volatility.

### 2.1 Setup

The model has three periods. At time 0, investors decide whether or not to pay a fixed cost  $c$  to become informed. If informed, they decide how to allocate their total attention  $K$  among the underlying risks. At time 1, informed investors receive signals about asset payoffs, and all investors submit their demands. At time 2, investors consume. Table 1 is a timeline of events in the model.

[Table 1 about here.]

### 2.1.1 No passive ownership

Without passive ownership, the model is similar to Admati (1985). The two key differences are (1) endogenous learning and (2) there are more risks than assets.

#### *Investors*

There are a unit mass of rational investors which fall into two groups: informed and uninformed. They both have CARA preferences over time 2 wealth. At time 1, informed investors receive signals about the assets' time two payoffs. The precision of these signals depends on how informed investors allocate their limited attention. Uninformed investors can only learn about terminal payoffs through prices. The third set of investors are noise traders, who have random demand at time 1, which prevents prices from being fully informative.

#### *Assets*

There are  $n$  assets, which I call stocks. Stock  $i$  has time 2 payoff:

$$z_i = a_i + f + \eta_i \tag{1}$$

where  $\eta_i \stackrel{\text{iid}}{\sim} N(0, \sigma_i^2)$  and  $f \sim N(0, \sigma^2)$ . In this economy there are  $n + 1$  risk-factors: one idiosyncratic risk-factor for each stock  $i$ ,  $\eta_i$ , and one systematic risk-factor,  $f$  that affects all stocks<sup>4</sup>. Each stock has  $\bar{x}_i$  shares outstanding and noise trader demand shocks  $x_i \stackrel{\text{iid}}{\sim} N(0, \sigma_{i,x}^2)$ . The  $\eta_i$ ,  $f$  and  $x_i$  shocks are jointly independent.

In the baseline version of the model, stocks are symmetric:  $\sigma_i^2 = \sigma^2$ ,  $\bar{x}_i = \bar{x}$  and,  $\sigma_{i,x}^2 = \sigma_x^2$ . This assumption is not needed, but it simplifies the intuition for the key learning trade-offs. For an extension where individual stocks load differently on systematic risk, and have heterogeneous volatility of their idiosyncratic risk-factors and supply shocks, see Section A.10.2 of the Appendix.

I also assume that the number of stocks  $n$  is sufficiently small so that idiosyncratic risk cannot be totally diversified away. Explicitly,  $\text{Var}(\frac{1}{n} \sum_{i=1}^n z_i) > \text{Var}(f)$ . This restriction to a small number of stocks is a reduced-form way of modeling transaction costs: Trading the first  $n$  stocks is free, but then trading costs go to infinity if the investor wanted to trade an

---

<sup>4</sup>There exists an equivalent economy where stock returns have the same correlation structure, but there is no systematic risk-factor. For example, suppose  $\text{cov}(\eta_i, \eta_j) = \sigma_f^2$  for all  $i$  and  $j$ . A systematic risk-factor, however, is needed to make the learning technology comparable between economies when passive ownership is and is not present. See Section A.12 of the Appendix for a discussion of this representation issue.

additional stock (see e.g. Merton (1987)). In the baseline calibration I set  $n = 8^5$ .

### *Signals*

If investor  $j$  decides to become informed, they receive noisy signals at time 1 about the payoffs of the underlying *stocks*:

$$s_{i,j} = a_i + (f + \epsilon_{f,j}) + (\eta_i + \epsilon_{i,j}) \quad (2)$$

where  $\epsilon_{i,j} \stackrel{\text{iid}}{\sim} N(0, \sigma_{\epsilon_{i,j}}^2)$ ,  $\epsilon_{f,j} \sim N(0, \sigma_{\epsilon_{f,j}}^2)$  and  $\epsilon_{i,j}$  are independent for all permutations of  $i$  and  $j$ , as well as independent from  $\epsilon_{f,j}$ . The signal noise,  $\epsilon$ , depends on how much attention investor  $j$  devotes to each *risk-factor* that affects the payoff of stock  $i$ :  $\eta_i$  and  $f$ . The learning technology governs how quickly signal noise decreases as more attention is devoted to a particular risk-factor.

### *Learning*

Investor  $j$  can allocate attention  $K_{i,j}$  to risk-factors  $\eta_i$  or  $f$  to reduce signal noise:

$$\sigma_{\epsilon_{i,j}}^2 = \frac{1}{\alpha + K_{i,j}}, \quad \sigma_{\epsilon_{f,j}}^2 = \frac{1}{\alpha + K_{n+1,j}} \quad (3)$$

where  $\alpha > 0$ . This differs from standard setups (e.g. Kacperczyk et al. (2016)), where the learning technology is  $\sigma_{\epsilon_{i,j}}^2 = \frac{1}{K_{i,j}}$ . In my setting,  $\sigma_{\epsilon_{i,j}}^2$  needs to be well defined even if an investor devotes no attention to risk-factor  $\eta_i$  or  $f$ <sup>6</sup>.  $\alpha$  can be viewed as informed investors having a “finger on the pulse” of the market. They know a little bit about each risk-factor, even without explicitly devoting attention to it. I set  $\alpha = 0.001$ , and discuss the sensitivity of the results to  $\alpha$  in Section A.11.5 of the Appendix.

Informed investors have a total attention constraint of  $\sum_i K_i \leq K$ . They also have a no forgetting constraint, so  $K_i > 0$  for all  $i$ . In the baseline calibration, learning capacity  $K$  is fixed to 1. Section A.10.1 of the Appendix discusses an alternative version of the model where investors can pay to increase learning capacity.

### *Portfolio Choice*

---

<sup>5</sup>Suppose there are 8 stocks,  $\sigma_f = 0.25$  and  $\sigma = 0.55$ . An equal-weighted portfolio of the 8 stocks would have a standard deviation of about 0.31, 25% larger than the standard deviation of the systematic risk-factor.

<sup>6</sup>This is because with more risks than assets, the risk-factors are not fully separable. For example, if  $\epsilon_{1,j}$  has infinite variance, but  $\epsilon_{f,j}$  has finite variance, the variance of  $s_{1,j}$  is still not well defined. In Kacperczyk et al. (2016), each of the rotated stocks is only exposed to one risk. Devoting no attention to any particular risk leads to a precision of zero, but this does not have spill-over effects on other stocks.

Define terminal wealth:

$$w_{2,j} = (w_{0,j} - \mathbb{1}_{informed,j}c) + \mathbf{q}'_j(\mathbf{z} - \mathbf{p}) \quad (4)$$

where  $w_{0,j}$  is initial wealth,  $c$  is the cost of becoming informed (in dollars),  $\mathbf{z}$  is the vector of terminal stocks payoffs,  $\mathbf{p}$  is the vector of time 1 prices and  $\mathbb{1}_{informed,j}$  is an indicator equal to 1 if investor  $j$  decides to become informed. Here, and everywhere else in the paper, boldface is used to denote vectors. The gross risk-free rate between time zero and time two is set to 1.

Investor  $j$  submits demand  $\mathbf{q}_j$  to maximize their time 1 objective function:

$$U_{1,j} = E_{1,j}[-exp(-\rho w_{2,j})] \quad (5)$$

where  $\rho$  is risk aversion.  $E_{t,j}$  denotes the expectation with respect to investor  $j$ 's time  $t$  information set. For informed investors, the time 1 information set is the vector of signals  $\mathbf{s}_j$  and the vector of prices,  $\mathbf{p}$ . For uninformed investors, the time 1 information set is just prices.

### *Prices*

Suppose we fix the share of informed investors, and the information choice of informed investors at some set of  $K_{i,j}$ 's. Then, the model is equivalent to Admati (1985). This is because investors do not independently receive information about the  $(n+1)^{th}$  risk-factor. Because there are more risks than signals/stocks, investors cannot rotate the economy (see e.g. Veldkamp (2011)) to think in terms of synthetic assets exposed only to risk-factor payoffs, rather than stock payoffs. This assumption is needed to solve the model using the closed form solutions in Admati (1985)<sup>7</sup>.

To understand the effect of thinking in terms of stock payoffs, consider the following example. Investor  $j$ 's stock 1 signal is:  $s_{1,j} = a_i + (f + \epsilon_{f,j}) + (\eta_1 + \epsilon_{1,j})$ . This is centered

---

<sup>7</sup>Without this assumption, there is no closed-form solution for the price function, as discussed in Section 6 of Admati (1985). To solve the model without this assumption, one would need to numerically solve for prices such that the market clears. The price function would be of the form  $p = \tilde{A}_0 + \tilde{A}_1\eta + \tilde{A}_2f + \tilde{A}_3\mathbf{x}$ , where  $\eta$  is the vector of stock-specific risk-factors and  $\mathbf{x}$  is a vector of supply shocks. It is difficult to solve for these  $A_i$  numerically, because one of the conditions for a solution includes the product of one of the price coefficients  $A_1$  with the inverse of another one of the price coefficients  $A_2^{-1}$ . This can lead to arbitrarily large offsetting entries in these matrices, and numerical instability.



on  $a_i + f + \eta_1$  so it is an unbiased signal about the payoff of stock 1. The variance of this signal is  $var(\epsilon_{f,j}) + var(\epsilon_{1,j})$  because all signal noise is independent. All investors know the correlation structure of stock returns, so when investor  $j$  is calculating a posterior mean for stock 2, they still consider the information in their signal for stock 1, as the stocks are correlated via their common exposure to systematic risk. Further, when deciding what to learn about, investors understand that devoting attention to systematic risk will reduce the variance of all of their stock signals.

To solve for prices, start by defining  $\mu$  as the vector of  $a_i$  i.e. the vector of mean stock payoffs. Further define  $\bar{x}$  as the vector of  $\bar{x}_i$  i.e. the vector of shares outstanding for each stock. Define the  $n \times (n + 1)$  matrix  $\Gamma$  as:

$$\Gamma = \begin{bmatrix} 1 & 0 & \dots & 0 & 1 \\ 0 & 1 & \dots & 0 & 1 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & 1 & 1 \end{bmatrix} \quad (6)$$

Defining  $\eta$  as a vector of  $\eta_i$ 's and  $f$  (where  $f$  is the last entry), terminal asset payoffs are  $\mathbf{z} = \mu + \Gamma\eta$ <sup>8</sup>

Define the variance of stock payoffs,  $V$  as:

$$V = \Gamma \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 & 0 \\ 0 & \sigma_2^2 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \sigma_n^2 & 0 \\ 0 & 0 & \dots & 0 & \sigma_f^2 \end{bmatrix} \Gamma' \quad (7)$$

---

<sup>8</sup>If the stocks had different loadings on systematic risk, the 1's in the last column would be replaced by  $\beta_i$ 's, i.e. the loadings of each stock on systematic risk, as discussed in the Appendix A.10.2.

Define the matrix of stock signal variances for investor  $j$  as:

$$S_j = \Gamma \begin{bmatrix} \frac{1}{\alpha+K_{1,j}} & 0 & \dots & 0 & 0 \\ 0 & \frac{1}{\alpha+K_{2,j}} & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \frac{1}{\alpha+K_{n,j}} & 0 \\ 0 & 0 & \dots & 0 & \frac{1}{\alpha+K_{n+1,j}} \end{bmatrix} \Gamma' \quad (8)$$

I assume all informed investors have the same attention allocation, so  $S_j = S$  for all  $j$  and  $K_{i,j} = K_i$ <sup>9</sup>.

Define the variance-covariance matrix of noise-trader shocks as  $U = \sigma_x^2 \mathbf{I}_n$  where  $\mathbf{I}_n$  is an  $n \times n$  identity matrix. Define the vector of realized noise-trader shocks as  $\mathbf{x}$ , which is normally distributed with mean zero and variance  $U$ . The available supply of each stock to informed and uninformed investors is  $\bar{\mathbf{x}} + \mathbf{x}$  i.e. the number of shares outstanding plus/minus demand from noise traders.

The equation for equilibrium prices comes directly from Admati (1985) (see Appendix A.1 for details):

$$\mathbf{p} = A_0 + A_1 \mathbf{z} - A_2 (\bar{\mathbf{x}} + \mathbf{x}) \quad (9)$$

### *Beliefs*

All investors extract an unbiased signal about stock payoffs from prices:

$$s_p = A_1^{-1} (p - A_0 + A_2 (\bar{\mathbf{x}} + \mathbf{x})) \quad (10)$$

Informed investors combine their signals  $s_{i,j}$ , with the information contained in prices  $s_p$  and update their prior beliefs using Bayes's law. Uninformed investors update their prior beliefs using only the information contained in prices.

### *Demands*

Demands are a function of private signals and prices. There are separate demand functions

---

<sup>9</sup>For discussions of non-symmetric equilibria, see e.g. Veldkamp (2011)

for the informed and uninformed:

$$\begin{aligned} \text{Uninformed: Demand} &= G_0 + G_{2,un} \mathbf{p} \\ \text{Informed, investor } j: \text{ Demand} &= G_0 + G_1 \mathbf{s}_j + G_{2,inf} \mathbf{p} \end{aligned} \tag{11}$$

where  $\mathbf{s}_j$  is the vector of signals received by investor  $j$  (see Appendix A.1 for details).

#### *Deciding to Become Informed*

At time zero, investor  $j$  decides whether or not to pay  $c$  and become informed. They make this decision to maximize the time 0 objective function:

$$U_{0,j} = -E_0[\ln(-U_{1,j})]/\rho \tag{12}$$

where the time 0 information set is the share of investors who decide to become informed. Equation 12 simplifies to:

$$U_{0,j} = E_0 [E_{1,j}[w_{2,j}] - 0.5\rho Var_{1,j}[w_{2,j}]] \tag{13}$$

because time two wealth is normally distributed. See Section A.9 of the Appendix for a discussion of how the preferences in Equation 12 differ from expected utility:  $U_{0,j} = E_{0,j}[U_{1,j}]$ .

### **2.1.2 Introducing passive ownership**

Passive ownership is modeled through an  $n + 1^{th}$  asset, which I call the ETF.

#### *Asset payoffs*

The ETF is only exposed to the systematic risk-factor  $f$  and has terminal payoff:

$$z_{n+1} = a_{n+1} + f \tag{14}$$

The ETF initially has average supply  $\bar{x} = 0$ , but is still subject to normally-distributed supply shocks  $x_{n+1}$ . These assumptions on the supply of the ETF are important for two reasons (1) Without supply shocks in the ETF, its price would be a fully revealing signal for the systematic risk-factor (2) the ETF must initially be in zero average supply so its introduction does not change the average quantity of systematic risk in the economy<sup>10</sup>.

---

<sup>10</sup>See Appendix A.2 for details.

*The size of passive ownership*

To model the growth of passive ownership, I introduce a new investor who can buy shares of the underlying stocks, and convert them into shares of the ETF. I assume that, unlike the atomistic informed and uninformed investors, this *ETF intermediary* is strategic: she understands that to create more shares of the ETF, she will have to buy more shares of the stocks, which will push up their *expected* prices. I emphasize *expected* because she still takes prices at  $t = 1$  as given. This is because I assume she can only submit a market order at  $t = 0$  i.e. she will have to decide how many shares of the ETF to create without knowing the  $t = 1$  prices of any security<sup>11</sup>.

Her objective function is the same as the objective function for the informed and uninformed investors:

$$U_{0,j} = E_0 [E_{1,j}[w_{2,int}] - 0.5\rho^i Var_{1,j}[w_{2,int}]] \quad (15)$$

where  $\rho^i$  is the intermediary's risk aversion. I assume that because assets 1 to  $n$  (the stocks) are symmetric, she must demand the same amount of each of them. If she buys  $v$  shares of every stock, this would take  $v \times n$  units of systematic risk out of the economy. To ensure that the amount of systematic risk in the economy is constant, I assume this allows her to create  $v \times n$  shares of the ETF. These assumptions imply that her only decision is how many shares of each stock to buy  $v$ .

Passive ownership is defined as  $v/\bar{x}_i$  i.e. the percent of each stock's shares outstanding which are part of the ETF. This maps almost exactly to the definition of passive ownership in the empirical exercises, which is the percent of each stock's shares outstanding owned by all passive funds.

With this technology, the intermediary's terminal wealth will be:

$$w_{2,int} = v \left( \underbrace{\sum_{i=1}^n (z_i - p_i)}_{\text{Idio. Risk}} - \underbrace{n(z_{n+1} - p_{n+1})}_{\text{Sys. Risk}} \right) \quad (16)$$

which is the average difference between the stocks' payoffs and their prices, minus the difference between the ETF's payoff and its price, scaled by how many shares she creates.

---

<sup>11</sup>For a more thorough discussion of these two assumptions, see Appendix A.3.

To create the ETF, the ETF intermediary is essentially stripping out the idiosyncratic risk from an equal-weighted basket of the stocks, and bearing it herself. She sells the systematic risk from this basket to informed and uninformed investors as an ETF. Having the intermediary bear this idiosyncratic risk is a reduced-form way of modeling basis risk that ETF arbitrageurs bear in the real world. While there can be no true basis risk in a model with no transaction costs and no price impact, I want to capture the risk inherent in creating shares of an ETF.

The optimal  $v$  mainly depends on  $\rho^i$ : if the intermediary is less risk averse, she will create more shares of the ETF. The increase in passive ownership over the past 30 years implies that  $\rho^i$  decreased. Given improvements in technology, trading speed, etc., it is reasonable to believe that ETF arbitrageurs are exposed to less risk now than they were in the past.

The size of passive ownership also depends on  $\rho$ ,  $\sigma$ ,  $\sigma_f$  and the share of informed investors. This is because these other parameters influence demand for the ETF, the ETF's price and thus the intermediary's profits. Section A.4 of the Appendix has more details on what determines the size of passive ownership.

#### *Signals and Learning Technology*

Informed investor  $j$  now receives signals about the payoffs of all the underlying assets, including a separate signal for the ETF:

$$\begin{aligned} s_{i,j} &= a_i + (f + \epsilon_{f,j}) + (\eta_i + \epsilon_{i,j}) \text{ for } i = 1, \dots, n \\ s_{n+1,j} &= a_i + (f + \epsilon_{f,j}) \end{aligned} \tag{17}$$

The learning technology and total attention constraint are unchanged from the economy where the ETF is not present.

#### *Price and Demands*

Having the intermediary submit a market order at  $t = 0$  means that the equilibrium price and demand functions are unchanged from the economy without passive ownership. Because this is a rational expectations equilibrium, all the investors anticipate the optimal  $v$ , given the model parameters. This means that informed and uninformed investors will treat the supply of each stock as  $\bar{x} - v$  and the supply of the ETF as  $n \times v$  when constructing their demand functions.

### 2.1.3 Relating ETFs in the model to ETFs in the real world

#### *ETFs vs. Futures Contracts and Index Mutual Funds*

In this economy the ETF looks like a futures contract: it is a claim, initially in zero net supply, on the payoff of the systematic risk-factor. Futures contracts, however, have existed for much longer than ETFs. If ETFs were equivalent to futures contracts, then we would not expect to see any of the empirical effects of growing ETF ownership (see e.g. Glosten et al. (2016), Ben-David et al. (2018)). The way the ETF is defined in this paper captures some features of the real-world, and misses others.

One thing it captures is that ETFs make it easier for investors to bet on systematic risk. This is consistent with the fact that ETFs are more divisible than futures, which allows more investors to trade them. For example, E-mini S&P 500 futures trade at around \$150,000 per contract, while SPY (the largest S&P 500 ETF) trades around \$300 per share (as of June 1, 2020). The investors who benefit from this increased divisibility are not just retirees trading in their 401K's. According to Daniel Gamba, former head of Blackrock's ETF business (iShares) *The majority of investors using ETFs are doing active management. Only about 30% of ETF investors look at these as passive funds...* (2016)<sup>12</sup>.

Another feature it captures is that ETFs have made it easier to hedge out/short systematic risk. According to Goldman Sachs Hedge Fund Monitor, "ETFs account for 27% of hedge funds short equity positions" (2016). This feature of the model is specific to the introduction of ETFs, relative to index mutual funds. Although index mutual funds existed before ETFs, (open-ended) mutual funds cannot be shorted.

#### *f as Sector-Specific Risk*

Another way to link the ETF in the model to the real world comes from viewing  $f$  as a sector-specific risk, rather than an economy-wide risk. ETFs cover more indexes and industries than futures contracts. These sector ETFs are popular: as of June 1, 2020, there was over \$170 Billion investment in State Street's 30 Sector ETFs. Another interpretation of the model is introducing an ETF that offers cheap diversification for particular industry. In Section A.13.1 of the Appendix I calibrate a version of the model to match the empirical effects of introducing sector ETFs in the late 1990's.

#### *ETF Creation/Redemption*

---

<sup>12</sup>Quoted in Balchunas (2016)

The model does not capture the creation/redemption mechanism of ETFs, an important feature that distinguishes them from index mutual funds and futures contracts. Other models like Cong and Xu (2016) have this feature. While this is an important channel, especially when talking about market-making in a Kyle (1985)-style model, I abstract away from this to focus on learning.

My simplified ETF creation technology does not exactly match the real world. ETF arbitrageurs do not hold on to the shares of the stocks they buy to create shares of the ETFs – they transfer them to an ETF custodian (e.g. State Street, BlackRock, Vanguard). This could be modeled by having the intermediary transfer the stocks she buys at  $t = 1$  to another (new) agent, an ETF custodian, who gives her shares of the ETF, which she sells immediately at  $t = 1$ . With this setup, the intermediary would have no asset holdings at  $t = 2$ .

With these alternative assumptions, all the qualitative results are unchanged. The quantitative difference is that creating shares of the ETF is less risky, so in equilibrium, the intermediary makes the ETF larger<sup>13</sup>. The reason I do not use this alternative setup is because with this ETF creation technology, the intermediary would be able to remove idiosyncratic risk from the economy by creating more shares of the ETF.

## 2.2 Equilibrium

At time 1, given  $K_i$ 's and the share of informed investors, the equilibrium is equivalent to that in Admati (1985): the demand functions ensure that the market clears, and beliefs formed using Bayes's law are rational. At time zero, an equilibrium requires: (1) no informed or uninformed investor would improve their expected utility by switching to the other type and (2) no informed investor would improve their expected utility by re-allocating their attention to different risk-factors. I use these two conditions to numerically solve the model<sup>14</sup>.

---

<sup>13</sup>In this scenario, the intermediary is only exposed to risk on her market order i.e. that the average prices of the stocks is higher than the price of the ETF due to positive realizations of stock-specific risk-factors or negative realizations of the stock-specific noise trader shocks.

<sup>14</sup>See Section A.8 of the Appendix for details

## 2.3 Learning Trade-offs

When an investor is deciding whether to devote attention to systematic or idiosyncratic risk, they face the following trade-off: (1) Learning about systematic risk leads to a more precise posterior belief about every asset (2) But, the volatility of systematic risk-factor ( $\sigma_f^2$ ) is low, relative to idiosyncratic risk-factors ( $\sigma^2$ ). This difference in volatilities means that there are more profit opportunities in the stock-specific risk factor than in the systematic risk factor. The ETF also affects this trade-off: If the ETF is not present, investors cannot take a bet purely on systematic risk, or idiosyncratic risks<sup>15</sup>.

To illustrate this trade-off, I present a few examples with only two stocks. Figure 2 shows the effect of learning on trading profits when there is no ETF and the assets are not exposed systematic risk i.e.  $z_i = a_i + \eta_i$ . Define excess trading profits as the difference between the profits of informed and uninformed investors in a particular security. The black line plots the excess profits of the informed investors in stock one, while the red line plots the excess profits of the informed investors in stock two. As we move to the right along the x-axis, informed investors are increasing their attention on stock 1. Initially, allocating more attention to stock one increases the informed investors' profit advantage in that stock, but eventually it hits a point of diminishing returns. The black line starts to slope down when the price becomes too informative about  $\eta_1$ . Because the stocks are symmetric, it is optimal for informed investors to allocate half their attention to each stock (vertical red line).

[Figure 2 about here.]

Compare this to Figure 3, where there are two stocks, but they are both exposed to a systematic risk-factor. Learning more about stock-specific risks (moving to the right along the x-axis) increases the informed investors' profit advantage, but eventually there are diminishing returns for two reasons. One reason is that prices become too informative, which is what also happened in the first example. The other reason is that both stocks are exposed to systematic risk, and informed investors are not learning much about a risk that affects

---

<sup>15</sup>Without the ETF, they cannot bet purely on an idiosyncratic risk, because they cannot perfectly hedge their exposure to systematic risk from holding that asset.



both stocks<sup>16</sup>. The slopes are different to the right/left of the optimum attention allocation (red vertical line) because the volatility of the systematic risk is lower than that of the stock-specific risks.

[Figure 3 about here.]

Finally, Figure 4 illustrates this learning trade off when there are two stocks, both exposed to systematic risk and idiosyncratic risks, and we introduce the ETF in zero average supply. Informed investors can now almost uniformly increase their profits in each stock by learning more about them. This is because they are able to take targeted bets on the stock-specific risk-factors by buying the stocks, and shorting the ETF. In equilibrium, informed investors learn more about stock-specific risks because there is more money to be made betting on  $\eta_i$ 's – the stock specific risk-factors are more volatile than the systematic risk-factor  $f$ . And because the investors are not very risk averse, with a CARA risk-aversion  $\rho$  of 0.1, they don't mind loading up on these volatile stock-specific risks.

[Figure 4 about here.]

## 2.4 Effects of passive ownership on learning

There are two learning margins: (1) How informed investors allocate their attention, which I call the intensive margin and (2) how many investors become informed, which I call the extensive margin. In this subsection, I walk through some examples to understand the effect of passive ownership on the intensive and extensive learning margins. These examples are not a calibration and are designed to illustrate the intuition.<sup>17</sup>

### 2.4.1 Intensive Learning Margin

Changing the share of informed agents affects which risks investors learn about in equilibrium. To shut off this channel, and isolate the intensive margin effects of passive ownership,

---

<sup>16</sup>Another factor is that without the ETF, informed investors cannot take targeted bets on the stocks without bearing some systematic risk. However, increasing attention on stock-one specific risk eventually has diminishes in Figure 2, where there is no systematic risk, which ensures this is not entirely driving the results in Figure 3.

<sup>17</sup>The parameters are mostly taken from Kacperczyk et al. (2016). See Appendix A.7 for details

I fix the share of informed agents. I compare attention to systematic risk across three scenarios: (1) *No ETF*, this is when investors cannot trade the ETF (2) *High  $\rho^i$* , this is when the investors have access to the ETF, but the intermediary is risk averse so it is in near zero supply (3) *Low  $\rho^i$* , investors can trade the ETF and the intermediary is closer to risk neutral, so the ETF is in positive supply.

Figure 5 shows how attention to systematic risk changes as we vary the size of passive ownership, fixing the share of informed agents at 60%. The left panel examines the effect of varying risk aversion  $\rho$ , fixing the volatility of the systematic risk-factor  $\sigma_f$  at 0.35. As risk aversion increases, informed investors devote more attention to systematic risk. The effects of increasing passive ownership, however, are ambiguous. If risk aversion is sufficiently low, passive ownership can *decrease* attention on systematic risk. If risk aversion is high, the opposite is true.

[Figure 5 about here.]

It seems counterintuitive that increases in passive ownership can lead to less learning about systematic risk. The mechanism driving this effect is what I call the *hedging* channel: the ETF allows informed investors to better isolate bets on stock-specific risk-factors. The larger the ETF is, the cheaper it is to hedge systematic risk.

The hedging channel will show up in investors' demand functions. For informed investors,  $G_1$  is a measure of how sensitive demand is to their private signals. Table 2 contains selected the entries of  $G_1$ . As with Figure 5, the share of informed investors is fixed at 60%. When the share of informed investors changes, all investors' posterior precision matrices change as well. This affects how aggressive investors are in betting on any signals and would confound the hedging channel effects<sup>18</sup>.

Because all the stocks have the same supply and have the same ex-ante risk,  $G_1$  is a symmetric matrix when the ETF is not present. The diagonal entries show how strongly investors react to signals about a particular stock. The off-diagonal entries show how investors hedge these bets. The diagonal entries of  $G_1$  are positive because when an investor gets a good signal about a stock, they buy more of it. The off-diagonal entries of  $G_1$  are negative

---

<sup>18</sup>This result is not unique to how informed investors respond to their own signals. Both informed and uninformed investors change their behavior in response to the signal contained in prices as well. For details, see Appendix A.11.4.

because they hedge these stock-specific bets by shorting an equal-weighted portfolio of all the other stocks.

For example, row 1 implies that a 1 unit higher signal about asset  $i$  leads to demand for 0.968 more shares of that stock, and this position is hedged by shorting -0.117 shares the other 7 stocks. This bet does not fully hedge out systematic risk, as 0.968 is greater than 7 times -0.117 (each stock has a unit loading on the systematic risk factor).

Compare this to the case where the ETF is present in zero average supply: Regardless of risk aversion, informed investors hedge out all the systematic risk embedded in each stock-specific bet with the ETF. Further, after introducing the ETF, informed investors bet *more* aggressively on the stocks with positive signals for low values of risk aversion/systematic risk.

[Table 2 about here.]

Figure 5 illustrates a key trade-off for informed investors in the model. If investors are risk averse, they care more about systematic risk because idiosyncratic risk can be diversified away. When we give them the ETF to trade on systematic risk directly, they want to learn even more about it. I call this the *market timing* channel. If investors are closer to risk neutral they care more about trading profits than risk. When you give them the ETF, it lets them take more targeted bets on volatile individual securities, and they learn more about the stock-specific risk-factors. This is one of the effects of the *hedging* channel.

The intensive margin effects also depend on the volatility of the systematic risk-factor  $\sigma_f$ . The right panel of Figure 5 examines the effect of varying  $\sigma_f$  fixing risk aversion  $\rho$  at 0.35. Increasing  $\sigma_f$  leads investors to devote more attention to systematic risk. This makes sense, because as a risk becomes more important to informed investors' terminal wealth, they allocate more attention to that risk. As with the left panel, however, the effect of passive ownership on attention allocation is ambiguous. If  $\sigma_f$  is sufficiently low, increasing passive ownership can lead to less learning about systematic risk, while if  $\sigma_f$  is sufficiently high, the opposite is true.

#### 2.4.2 Extensive Learning Margin

To examine the extensive margin effects of passive ownership, I fix the cost of becoming informed  $c$ , and compare how many investors become informed across the same three scenar-

ios: (1) no ETF (2) high  $\rho^i$  and (3) low  $\rho^i$ . Figure 6 shows the relationship between the cost of becoming informed (in dollars) and the percent of rational investors who decide to become informed in equilibrium. The *risk-bearing capacity* of the economy depends jointly on the share of informed agents, the volatility of the systematic risk factor and informed/uninformed investors' risk aversion<sup>19</sup>.

The left panel presents a scenario where risk aversion and the volatility of the systematic risk-factor are high, so the risk-bearing capacity of the economy is low. As we increase the cost of becoming informed, fewer investors become informed. As we increase the size of passive ownership, fewer investors become informed. This is because when the economy has a low risk-bearing capacity, the ETF makes the uninformed investors relatively better off. I call this the *diversification* channel.

The right panel presents a scenario where risk aversion and the volatility of the systematic risk-factor are low, so the risk-bearing capacity of the economy is high. With these parameters, as passive ownership increases, more investors become informed. This is because almost risk-neutral investors are willing to bet aggressively on signals about stock-specific risks. This is another effect of the *hedging* channel: Passive ownership increases the benefit of being informed.

[Figure 6 about here.]

Figures 5 and 6 show that the intensive and extensive margin effects of increasing passive ownership are ambiguous. This is because of the three competing channels outlined above: The hedging channel leads to more investors becoming informed, and increases the share of attention allocated to stock-specific risks. The market timing channel leads investors to devote more attention to systematic risk. Finally, the diversification channel leads fewer investors to become informed.

The natural next step is to calibrate the model to the data, and understand which of these competing effects dominates. It is difficult, however, to empirically observe how many

---

<sup>19</sup>The concept of *risk-bearing capacity* is designed to capture the following intuition: It is possible to change one of (1) risk aversion  $\rho$ , (2) the volatility of the systematic risk factor  $\sigma_n$ , or (3) the share of informed agents, and offset the effect of this change on the intensive/extensive learning margins by varying the other two. For example, consider an increase in  $\rho$ . This would tend to decrease the share of informed agents, and increase attention on systematic risk. It is possible, however, to keep learning mostly the same by decreasing  $\sigma_n$  and/or increasing the share of informed agents.

investors are informed and which risks investors are learning about. In the next subsection, I develop measures of price informativeness that are easily observable in the data.

## 2.5 Effects of passive ownership on price informativeness

In this subsection, I quantify the effect of increasing passive ownership on price informativeness. The natural first step is to derive a model-based measure of price informativeness at  $t = 1$ . The issue is that there is no consensus on the right way to theoretically measure price informativeness, and many price informativeness measures are hard to map to the data.

For example, Grossman and Stiglitz (1980) defines price informativeness as a *conditional* covariance, which requires identifying the right set of conditioning variables, which academic economists still disagree on. Based on Grossman-Stiglitz, Bai et al. (2016) measure price informativeness as the variance of fundamentals, conditional on prices. Dávila and Parlatore (2019) measure price informativeness as the variance of prices, conditional on fundamentals, effectively switching the left-hand-side and right-hand-side variables of the main regression in Bai et al. (2016). The correct measure of future fundamentals is also not obvious<sup>20</sup>.

Instead, I focus on the observable variables discussed in the introduction: trading volume, returns and volatility. I create model analogues of these objects, and simulate the economy to determine the effect of growing passive ownership these alternative measures of price informativeness. To map the model to the stylized facts, I label  $t = 1$  as the pre-earnings announcement date, and  $t = 2$  as the earnings announcement.

### *Pre-Earnings Trading Volume*

Although the model features a continuum of investors, when simulating the economy, there are a finite number, which I set to 10,000. At  $t = 0$ , I assume all of the investors are endowed with  $1/10,000^{th}$  of  $\bar{x}$ . One way to define trading volume is the difference between investors' initial holdings, and their holdings after markets clear. This measure, however, would be contaminated by the noise trader shock. To account for this, I measure trading volume as the difference between initial holdings and final holdings, divided by the total supply of the asset, which *includes* the noise shock.

---

<sup>20</sup>In a static model like Grossman-Stiglitz, there are a finite number of cash flows, but in reality, firms are long-lived. Maybe fundamentals are *all* futures cashflows, which are hard to measure. It is also not clear if earnings are the right measure of future fundamentals as management has some control over earnings growth (see e.g. Schipper (1989)).

Let  $J$  denote the total number of investors. Then pre-earnings volume is defined as:

$$\sum_j^J |\mathbf{q}_j - (\bar{\mathbf{x}} + \mathbf{x}) / (J)| \quad (18)$$

where the first term  $\mathbf{q}_j$  is investor  $j$ 's demand, and the second term  $(\bar{\mathbf{x}} + \mathbf{x}) / (J)$  is investor  $j$ 's share of the initial endowment  $\bar{\mathbf{x}}$ , adjusting for the noise shock  $\mathbf{x}$ .

There are two main factors that affect trading volume in the model: (1) The share of investors who decide to become informed. As more investors become informed, there are more different signals in the economy, and thus more trading. Uninformed investors all submit the same demand, so in the limit with all uninformed investors, there would be no trading volume after accounting for the noise shock (2) Attention allocation. As more attention is devoted to the individual stocks, informed investors have more precise posterior beliefs, and are more willing to bet more aggressively on their signals. Less trading volume is therefore evidence of fewer informed traders, and less learning about stock-specific risks.

#### *Pre-Earnings Drift*

Define the pre-earnings drift as:

$$DM = \begin{cases} \frac{1+r_{(0,1)}}{1+r_{(0,2)}} & \text{if } r_2 > 0 \\ \frac{1+r_{(0,2)}}{1+r_{(0,1)}} & \text{if } r_2 < 0 \end{cases} \quad (19)$$

where  $r_{(0,t)}$  is the cumulative market-adjusted return from 0 to  $t$ <sup>21</sup>. The pre-earnings drift will be near one when the return at  $t = 2$  is small relative to the return at  $t = 1$ .  $DM_{i,t}$  will be less than one when the  $t = 2$  return is large, relative to the returns at  $t = 1$ . If  $r_2$  is negative, this relationship would be reversed, which is why the measure is inverted when  $r_2$  is less than zero<sup>22</sup>. To compute this measure, I save the prices at  $t = 0$ ,  $t = 1$  (calculated using the equilibrium in Admati (1985)),  $t = 2$  (terminal payoffs), and compute returns as  $r_{(t-n,t)} = \frac{p_t - p_{t-n}}{p_{t-n}}$  and  $r_t = \frac{p_t - p_{t-1}}{p_{t-1}}$ . Higher drift implies more informative prices.

---

<sup>21</sup>I work with market-adjusted returns to account for the effect of growing passive ownership on risk premia. Market-adjusted returns are defined as the return of the stock minus the average return of all stocks, to make things comparable between the scenario when the ETF is and is not present. See Section A.11.2 of the Appendix for details.

<sup>22</sup>This is similar to the price jump ratio in Weller (2017), but can be computed for all stocks. Weller has to filter out over 50% of earnings announcements because the denominator of his measure can be near zero.

### *Share of Volatility on Earnings Days*

Define the share of volatility on earnings days as:

$$\frac{r_2^2}{r_1^2 + r_2^2} \quad (20)$$

If prices are not informative before earnings announcements, we would expect earnings day volatility to be large, relative to total volatility<sup>23</sup>.

### *Effect of Learning on Price Informativeness*

It seems natural that price informativeness should be related to the intensive and extensive learning margins. To measure the intensive margin's effect on price informativeness, I run the following regression:

$$\text{Price Informativeness} = \gamma + \beta \text{Attention to Idio. Risk} + \text{Fixed Effects} + \text{Error} \quad (21)$$

I simulate the economy for all values of  $\rho$  and  $\sigma_n$  between 0.05 and 0.45, in increments of 0.1. For each of these choices of  $\rho$  and  $\sigma_n$  I simulate the economy for all values of the share informed agents between 0.2 and 0.7, in increments of 0.05. The unit of observation is the average price informativeness for a particular set of  $\rho$ ,  $\sigma_n$  and the share of informed agents across 10,000 simulations.

Panel A of Table 3 contains the results. I include fixed effects for the share of informed agents to rule out extensive margin effects. As expected, increased attention to stock-specific risks leads to increased pre-earnings drift, decreased earnings-day volatility and increased pre-earnings trading volume.

I run a similar regression for the extensive margin's effect on price informativeness:

$$\text{Price Informativeness} = \gamma + \beta \text{Share Informed} + \phi \text{Attn. to Idio. Risk} + \text{FE} + \text{Error} \quad (22)$$

The unit of observation is the same as in 21. Panel B of Table 3 contains the results. I include fixed effects for  $\rho$  and  $\sigma_n$ , and control for informed investors' attention to idiosyncratic risk to rule out intensive margin effects. As expected, increasing the share of informed agents leads to increased pre-earnings drift, decreased earnings-day volatility and increased pre-earnings

---

<sup>23</sup>These results are not sensitive to using squared returns i.e. focusing on extreme observations. I find similar results working with absolute returns e.g.  $|r_2|/(|r_1| + |r_2|)$ .

trading volume.

[Table 3 about here.]

Table 3 confirms that the intensive and extensive learning margins drive changes in price informativeness. The growth of passive ownership affects both of these learning margins, so it should also have an effect of price informativeness. In the next section, I calibrate a version of the model to match the empirical rise of passive ownership.

*Calibration to the rise of passive ownership*

Between 1990 and 2018, passive ownership grew nothing to owning 15% of the US stock market. In Figure 7, I compare two scenarios: (1) No ETF (2) ETF owning 15% of each stock. To allow for both intensive and extensive margin learning effects, I fix the cost of becoming informed to match a particular share of informed agents when the ETF is not present. Then, I calculate how many agents optimally become informed in equilibrium at this cost when the ETF owns 15% of the market. All the price informativeness measures are only calculated for the stocks i.e. assets 1 to  $n$ .

[Figure 7 about here.]

The top 3 panels are averages of the price informativeness measures in an economy with low risk aversion  $\rho = 0.05$  and low volatility of the systematic risk factor  $\sigma_f = 0.05$ . Increasing the share of informed agents (moving to the right along the x-axis) unambiguously increases price informativeness: it increases the pre-earnings drift, decreases earnings-day volatility and increases pre-earnings trading volume. Increasing the size of passive ownership decreases the pre-earnings drift, suggesting less informative prices. However, passive ownership also decreases the volatility on earnings announcement dates, and increases pre-earnings trading volume suggesting more informative prices. This is evidence of an economy where the hedging channel can dominate the diversification and market timing channels.

The bottom 3 panels are averages of the price informativeness measures in an economy with high risk aversion  $\rho = 0.35$  and high volatility of systematic risk  $\sigma_f = 0.35$ . All three measures suggest that growing passive ownership leads to less informative prices in this economy. The effect of passive ownership is stronger when the share of informed agents is larger because there is more room for the extensive learning margin to work. This is evidence



of an economy where the hedging channel is dominated by the diversification and market timing channels.

As with the extensive and intensive learning margins, the effect of passive ownership on the price informativeness measures is ambiguous. In the next section, I map the model-based measures of price informativeness to the data, so I can test which effects dominate empirically.

### 3 Mapping the model to the data

In this section, I construct an empirical measure of passive ownership that matches the definition of passive ownership in the model. I also define empirical analogues of the three model-based measures of price informativeness using trading volume, returns and volatility around earnings announcements. The cross-sectional average of all three price informativeness measures has declined over the past 30 years.

#### 3.1 Defining passive ownership

Passive funds are defined as all index funds, all ETFs, and all mutual funds with “index” in the name. Index funds are identified using the index fund flag in the CRSP mutual fund data<sup>24</sup>. Passive ownership is defined as the percent of a stock’s shares outstanding owned by passive funds. This maps almost exactly to the definition of passive ownership in the model  $v/\bar{x}$ , the percent of each stock’s shares outstanding owned by the ETF.

To calculate passive ownership, I need to identify the holdings of passive funds, which I obtain from the Thompson S12 data. I use the WRDS MF LINKS database to connect the funds identified as passive in CRSP with the S12 data. If a security never appears in the S12 data, I assume its passive ownership is zero. Figure 8 shows that passive ownership increased from almost zero in 1990, to now owning about 15% of the US stock market. As of 2018, passive ownership was over 40% of total mutual fund and ETF assets.

[Figure 8 about here.]

---

<sup>24</sup>Over 90% of the fund-quarter observations I identify as passive have a non-missing index fund flag in CRSP. The other 10% are exclusively identified by the name matching and/or the ETF flag in CRSP.

I believe this is a conservative definition of passive ownership, as there are institutional investors which track broad market indexes, but are not classified as mutual funds, and thus do not appear 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 is also omitted in my definition of passive management.

## 3.2 Data for Constructing Price Informativeness Measures

All return and daily volume data are from CRSP. I restrict to ordinary common shares (share codes 10 and 11) traded on major exchanges (exchange codes 1 to 3). I merge CRSP to I/B/E/S (IBES) using the WRDS linking suite. I use the earnings release times in IBES to identify the first date investors could trade on earnings information during normal market hours. If earnings are released before 4:00 PM eastern time between Monday and Friday, that day will be labeled as the effective earnings date. If earnings are released on or after 4:00 PM eastern time between Monday and Thursday, the next day will be labeled as the effective earnings date (as long as the next day is not a trading holiday). If earnings are released on Friday after 4:00 PM eastern, over the weekend, or on a trading holiday, the next trading date is labeled as the effective earnings date.

I define quarterly earnings per share as the “value” variable from the IBES unadjusted detail file.<sup>25</sup> All other firm fundamental information is from Compustat. Total institutional ownership is the percent of a stock’s shares outstanding held by all 13-F filing institutions. Institutional ownership is merged to CRSP on CUSIP, or historical CUSIP. If a CUSIP never appears in the 13-F data, institutional ownership is assumed to be zero.

## 3.3 Measure 1: Pre-earnings volume

Using pre-earnings trading volume to quantify price informativeness is motivated by the literature on asymmetric information (Akerlof (1978), Milgrom and Stokey (1982)). As information asymmetries become larger, uninformed agents are less willing to trade because of adverse selection. They are concerned that the only people willing to trade with them are better informed, so any trades they make are guaranteed to be bad deals. In the stock

---

<sup>25</sup>All results are similar when using Diluted Earnings Per Share Excluding Extraordinary Items (EPSFXQ) in Compustat.

market, an uninformed investor may prefer to delay trading until uncertainty is resolved (see e.g., Admati and Pfleiderer (1988), Wang (1994)).

In the model, however, adverse selection does not drive the relationship between price informativeness and trading volume. Trade is generated by differences of opinion, which are amplified by the precision of investors' beliefs. Therefore, there is less trading before the earnings announcement date ( $t = 1$ ) when fewer investors become informed, and when investors learn less about stock-specific risks. Prices aggregate information, so if there are fewer different opinions or investors trade less aggressively on their opinions, prices are less informative.

Unlike in the model, there are many dates between earnings announcements. The model's predicted drop in trading volume may be spread out over the month (22 trading days) before an earnings announcement. Let  $t$  denote an effective earnings announcement date. Define abnormal volume for firm  $i$ , from time  $t - 22$  to  $t + 22$  as:

$$AV_{i,t+\tau} = \frac{V_{i,t+\tau}}{\overline{V_{i,t-22}}} = \frac{V_{i,t+\tau}}{\sum_{k=1}^{63} V_{i,t-22-k}/63} \quad (23)$$

Where abnormal volume,  $AV_{i,t+\tau}$ , is volume divided by the historical average volume for that firm over the past quarter<sup>26</sup> In Equation 23,  $V_{i,t+\tau}$  is total daily volume for stock  $i$  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:

$$AV_{i,t+\tau} = \alpha + \sum_{k=-21}^{22} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau} \quad (24)$$

The right-hand side variables of interest are a set of indicators for days relative to the earnings announcement. For example,  $\mathbf{1}_{\{\tau=-15\}}$  is equal to one 15 trading days before the nearest earnings announcement, and zero otherwise. The regression includes all stocks that can be matched between CRSP and IBES, and a  $\pm 22$  day window around each earnings announcement.

I run this regression for three sample periods: (1) 1990-1999 (2) 2000-2009 (3) 2010-2018.

---

<sup>26</sup>All results are robust to instead using the average volume for that firm over the past year.

Figure 9 plots the estimates of  $\beta_k$  for  $k = -21$  to  $k = -2$ . For each day, the average abnormal volume is statistically significantly lower in the third period, relative to the first period.

[Figure 9 about here.]

Figure 9 confirms that there has been a drop in trading volume throughout the month before each earnings announcement. Define cumulative abnormal pre-earnings volume as:

$$CAV_{i,t} = \sum_{\tau=-22}^{-1} AV_{i,t+\tau} \quad (25)$$

the sum of abnormal trading volume from  $t - 22$  to  $t - 1$  for firm  $i$  around earnings date  $t$ .  $CAV_{i,t}$  is one of my main empirical measures of price informativeness. Lower values of  $CAV_{i,t}$  translate to less pre-earnings trading and are evidence of less informative prices.

Figure 10 shows the value-weighted average of  $CAV_{i,t}$  by year. Between the 1990's and 2010's, average  $CAV_{i,t}$  declined by about 1. This can be interpreted as a loss of about 1 trading-day's worth of volume over the 22-day window before earnings announcements.

[Figure 10 about here.]

### 3.4 Measure 2: Pre-earnings drift

The pre-earnings drift i.e., the fact that firms with strong (weak) earnings tend to have positive (negative) pre-earnings returns has been studied extensively (see e.g., Ball and Brown (1968), Foster et al. (1984), Weller (2017)). If investors are trading on signals of good news before earnings are released, or the firm gives guidance of strong future performance, we expect prices to increase before the earnings announcement date.

This also happens in the model. Suppose one of the stocks is going to have a high payoff at  $t = 2$ . As we increase the share of informed investors, or we increase informed investors' attention on stock-specific risk-factors, the price at  $t = 1$  will be relatively higher. Empirically, this upward drift may happen over the month before the earnings announcement as informed investors want to avoid moving the market against them (see e.g. Kyle (1985)).

Let  $E_{i,t}$  denote earnings per share for firm  $i$  in quarter  $t$  in the IBES Unadjusted Detail

File<sup>27</sup>. Define standardized unexpected earnings (SUE) as the year-over-year (YOY) change in earnings, divided by the standard deviation of YOY changes in earnings over the past 8 quarters.

$$SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})} \quad (26)$$

Define market-adjusted returns,  $r_{i,t}$ , as in Campbell et al. (2001): the difference between firm  $i$ 's excess return and the excess return on the market factor from Ken French's data library.

Each quarter, I sort firms into deciles of  $SUE$ , and calculate the cumulative market-adjusted returns over the 22 trading days prior to the earnings announcement. Figure 11 shows the average pre-earnings cumulative returns by SUE decile for two different time periods: 2001-2007 and 2010-2018. The 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<sup>28</sup>. 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 2018, firms in each decile move less before earnings days than between 2001 and 2007.

[Figure 11 about here.]

Figure 11's apparent decline in the pre-earnings drift could be driven by differences in overall return volatility or average returns between the two time periods. The drift magnitude variable from the model, however, is designed to capture the *share* of earnings information incorporated into prices before the announcement date. Define the pre-earnings drift for firm  $i$  as the cumulative market-adjusted gross return from  $t - 22$  to  $t - 1$ , divided by the cumulative returns from  $t - 22$  to  $t$ , where  $t$  is an earnings announcement:

$$DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases} \quad (27)$$

---

<sup>27</sup>All results are robust to using earnings per share in Compustat. I work with IBES earnings to be consistent with analyst estimates, which I discuss in Section 4.6

<sup>28</sup>For a detailed discussion of how Reg FD affects my results, see Section C.2.1 of the Appendix.

The pre-earnings drift will be near one when the earnings day move is small relative to cumulative pre-earnings returns.  $DM_{i,t}$  will be less than one when the earning-day return is large, relative to the returns over the previous 22 days. If  $r_t$  is negative, this relationship would be reversed, which is why the measure is inverted when  $r_t$  is less than zero.<sup>29</sup> I work with gross returns, rather than net returns, to avoid dividing by numbers near zero<sup>30</sup>

$DM_{i,t}$  is going to be my second main empirical measure of pre-earnings price informativeness. Lower values of  $DM_{i,t}$  imply less information is getting into prices before earnings announcement dates. Figure 12 shows the cross-sectional average value of  $DM_{i,t}$  by year. The pre-earnings drift decreased by about -0.02 between 1990 and 2018.

[Figure 12 about here.]

### 3.5 Measure 3: Earnings-days' share of annual volatility

The last two subsections showed there is less trading before earnings announcements, and the pre-earnings drift declined. If the total amount of information is not changing over time, we would expect there to be larger returns on earnings days, relative to all other days. In the model, when fewer investors become informed, or investors devote less attention to stock-specific risk factors, earnings day returns ( $t = 2$ ) become more volatile, relative to non-earnings days ( $t = 1$ ). The empirical analogue of this is the share of total annual volatility occurring on earnings dates.

Specifically, define the quadratic variation share (QVS) for firm  $i$  in year  $t$  as:

$$QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2 \quad (28)$$

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  $t$ .  $QVS$  is going to be my third main empirical measure

<sup>29</sup>Section B.3 of the Appendix presents an alternative definition of the pre-earnings drift using squared returns and further motivates my specification for  $DM_{i,t}$ .

<sup>30</sup>It has been well documented (see e.g. McLean and Pontiff (2016)) that the post-earnings drift has declined. To ensure my results are not driven by this trend, I calculate alternative measures of the pre-earnings drift replacing  $1 + r_{(t-22,t)}$  with  $1 + r_{(t-22,t+n)}$  for  $n$  between 1 and 5. All my results are qualitatively and quantitative unchanged using these alternative pre-earnings drift measures.

of price informativeness. If relatively more information is being learned and incorporated into prices on earnings dates, we would expect larger values of  $QVS$ .

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 13 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 IBES. Average  $QVS$  increased from 3.0% in 1990 to almost 16% in 2018. Figure 24 in the Appendix shows that the increase in  $QVS$  was due to a simultaneous increase in the numerator (volatility on earnings days) and a decrease in the denominator (volatility on all other days).

[Figure 13 about here.]

### 3.6 Robustness of Stylized Facts

These downward trends in market efficiency could be unrelated to the information released on earnings days. To rule this out, I propose the following placebo test: choose the date 22 trading days before each earnings announcement to be a placebo earnings date. I then reconstruct the time-series averages of the pre-earnings volume, drift and share of volatility on these placebo earnings days. In Section 3.6 of the Appendix, Figure 28 shows that there is no drop in volume before the placebo earnings dates in the last third of the sample. Figure 29 shows that there is no downward trend in the pre-earnings drift for the placebo earnings dates. Figure 33 shows there is no upward trend in the share of volatility on the placebo earnings dates. These figures are similar if you use randomly selected dates as placebo earnings announcements. These results confirm that the changes in price informativeness are specific to earnings days.

As an additional check, Section B.6.2 of the Appendix examines volume, drift and volatility around Federal Open Market Committee (FOMC) meeting dates instead of placebo earnings dates. The growth of index funds and ETFs has made it easier to trade on systematic information. It is possible that investors now focus on gathering information about systematic risks, so stock prices are more informative about systematic news, at the expense of firm-specific news. I find no trend toward decreased, or increased, efficiency in the incorporation FOMC meeting information at the stock-level. This also confirms that the reduction in efficiency only applies to firm-specific information.

## 4 Reduced-form relationship between passive ownership and price informativeness

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 calibrate the model to qualitatively match the reduced-form results. Finally, I present evidence for the model’s learning mechanisms via the correlation between passive ownership and decreased information gathering.

### 4.1 Pre-earnings volume

I run the following regression with quarterly data to measure the relationship between pre-earnings volume and passive ownership:

$$\Delta CAV_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \quad (29)$$

where cumulative abnormal pre-earnings volume,  $CAV_{i,t}$ , is defined in Equation 25.  $\Delta$  is a year-over-year change, matching on fiscal quarter<sup>31</sup>. 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  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . 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 prevent a firm-size effect driving my results. I also include firm and year/quarter fixed-effects. The time fixed effects ensure I am comparing firms at the same point in time with different increases in passive ownership. This allays concerns that my results are driven by simultaneous trends in passive ownership and price informativeness.

---

<sup>31</sup>One concern with the measure of abnormal volume in Equation 23 is that only using historical data from the previous quarter would lead to seasonal patterns. Using the year-over-year change, and matching on fiscal quarter should alleviate concerns that seasonality is driving my results, as I am comparing the same season within a firm over time. In addition, I have replicated all my results defining abnormal volume as volume relative to average volume over the past year, and find it has no qualitative effect.



Standard errors are computed using panel Newey-West with 8 lags<sup>32</sup>.

Table 4 contains the regression results. The coefficient on  $\Delta Passive_{i,t}$  in the value-weighted specification with all controls/fixed effects (column 3) implies that a 15% increase in passive ownership would lead to a decline in cumulative abnormal pre-earnings volume of -3.6. *CAV* (level) has a value-weighted mean of 22.6 and a standard deviation of 10.4. So, this decline of -3.6 is about 1/3 of a standard deviation.

[Table 4 about here.]

## 4.2 Pre-earnings drift

I run the following regression with quarterly data to measure the relationship between the pre-earnings drift and passive ownership:

$$\Delta DM_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \quad (30)$$

where  $DM_{i,t}$  is defined as in Equation 27. Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . Fixed effects are year/quarter and firm. Standard errors are computed using panel Newey-West with 8 lags.

The regression results are in Table 5. The coefficient on  $\Delta Passive_{i,t}$  in the value-weighted specification with all controls/fixed effects (column 3) implies that a 15% increase in passive ownership would decrease the pre-earnings drift by -0.0145. *DM* (level) has a value-weighted mean of 0.971 and a standard deviation of 0.033. So this decline of -0.0145 is about 1/2 of a standard deviation.

[Table 5 about here.]

---

<sup>32</sup>All results are robust to double-clustering standard errors at the firm/year level.

### 4.3 Earnings-days' share of annual volatility

I run the following regression with annual data to measure the relationship between changes in earnings day share of annual volatility, and changes in passive ownership:

$$\Delta QVS_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \quad (31)$$

where  $QVS$  is defined in Equation 28. Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . Fixed effects are year/quarter and firm. Standard errors are computed using panel Newey-West with 2 lags<sup>33</sup>

The regression results are in Table 6. The coefficients on  $\Delta Passive_{i,t}$  in the value-weighted specification with all controls/fixed effects (column 3) implies that a 15% increase in passive ownership would lead to an increase in earnings-day volatility of 0.057.  $QVS$  (level) has a value-weighted mean of 0.085 and a standard deviation of 0.11. So this increase of 0.057 is about 1/2 of a standard deviation.

[Table 6 about here.]

### 4.4 Robustness of Reduced-Form Results

To confirm that my results are specific to earnings days, I perform two placebo tests. The first set of placebo earnings dates are 22 trading days before each earnings announcement. The second are all scheduled FOMC meetings. Appendix Table 20 compares the original regression results to the placebo results in the specifications with all controls and firm-fixed effects.

I focus on the earnings-day share of volatility in these tests, as I believe it offers the cleanest comparison. For the FOMC announcement dates, looking at year-over-year changes for the  $n^{th}$  annual announcement does not make much sense, as there is (1) no analogue of fiscal year to account for seasonality and (2) they do not occur at the same time every year. The first point also applies to the days between earnings announcements. All of the

---

<sup>33</sup>I use two lags, instead of 8 as in the previous two sub-sections, because this regression is run at the annual frequency.

placebo results are insignificant, confirming that the relationship between passive ownership and volatility are all specific to earnings days. As an additional check, 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  $\Delta Passive$ .

Two threats to identification are (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)). Section C.2.1 of the Appendix shows that all my results are robust to only using data after Reg FD passed. Section C.2.2 of the Appendix shows that my results are robust to controlling for the AT measures in Weller (2017). Section C.3 of the Appendix discusses possible omitted variables in the baseline regressions. It is not possible to discuss every alternative hypothesis, so outside of explicitly testing these 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.

## 4.5 Calibrating the model to match the reduced-form results

In the model, passive ownership has an ambiguous effect on price informativeness. Equating increasing the size of the ETF in the model to the increases in passive ownership in the data, I can calibrate the model to match the empirical results. I want the calibration to satisfy two conditions. First, passive ownership *quantitatively* matches the data at 15%. Second, price informativeness monotonically decreases after (1) introducing the ETF in zero average supply and (2) growing the ETF to 15% of the market. To this end, I search over all possible values of (1) the share of informed investors when the ETF is not present (2) risk aversion  $\rho$  (3) the volatility of the systematic risk factor  $\sigma_f$  (4) risk aversion of the ETF intermediary  $\rho^i$ .

The results are in Table 7. Pre-earnings volume and the pre-earnings drift monotonically decrease as passive ownership increases, while the share of volatility on earnings days monotonically increases. The cost of becoming informed is set so 60% learn in equilibrium when the ETF is not present. At this cost, 50% learn when the ETF is 15% of the market. This is evidence of the extensive margin effect at work. Risk aversion  $\rho = 0.15$ , and the volatility of the systematic risk-factor  $\sigma_f^2=0.25$ . To match the average passive ownership of 15% with

all these other parameters,  $\rho^i$  is set to 1.

The row labeled data is the effect of a 15% increase in passive ownership based on value-weighted reduced-form estimates. Although I am able to qualitatively match the patterns here, the match is not quantitatively strong. Although drift and volatility are within an order of magnitude, volume is way off. This is probably due to the fact that there is no notion of “abnormal” volume in the model. There is no trading other than at  $t = 1$  so there is no reference point to compare this to.

[Table 7 about here.]

The model can qualitatively match the empirical results for the three measures of price informativeness. In the next subsection, I test the model’s predictions for the effect of rising passive ownership on information gathering.

## 4.6 Effect of passive ownership on information gathering

The model shows that if risk aversion  $\rho$  or volatility of the systematic risk factor  $\sigma_n$  are sufficiently high, passive ownership can decrease the share of informed investors and decrease investors’ attention on stock-specific risks.

This is consistent with the intuition 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 analyst coverage/accuracy on passive ownership:

$$\Delta Outcome_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \quad (32)$$

Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . Fixed effects include year/quarter and firm.

In Equation 32, 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* (3) the average time (in months) between analyst updates. 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. After excluding these observations, accuracy is Winsorized at the 1% and 99% level each year. All results are robust to normalizing by the stock price instead of the consensus forecast.

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.

Another 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:

$$\Delta DL_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \quad (33)$$

where  $\Delta$  is the change from year  $t - 1$  to year  $t$ .  $DL_{i,t}$  is the number of non-robot downloads of 10-K's, 10-Q's and 8-K's in the 22 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 robots may automatically download all filings at release, or update a database periodically. These robot downloads do not coincide with the intuition of information gathering for immediate trading<sup>34</sup>. 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, so the trend would bias me against finding any results.

---

<sup>34</sup>It is possible that quantitatively-driven investment firms, which use the information in downloaded SEC filings when developing trading strategies, are classified as robots. I conjecture that these firms download the information as it is released i.e. on the earnings day itself. They may use this information right away, or throughout the whole quarter, so the actual timing of the download would be independent of when it is used for informed trading. I do not think that these firms are re-downloading the same SEC filings every time they trade.

Table 8 contains the regression results. Consistent with decreased information gathering, increases in passive ownership are correlated with the fewer analysts covering a stock, lower analyst accuracy, less frequent analyst updates. Firms with increases in passive ownership also experience decreases in pre-earnings downloads of SEC filings.

[Table 8 about here.]

Given the negative correlation between passive ownership and information gathering, raises the possibility that the reduced-form results are driven by reverse causality. Maybe passive ownership happened to increase the most in stocks that had the biggest decline in passive ownership. In the next section, I exploit quasi-exogenous increases in passive ownership to rule out this alternative.

## **5 Effect of quasi-exogenous increase in passive ownership on price informativeness**

In this section, I exploit S&P 500 index additions, as well as Russell 1000/2000 reconstitutions to identify increases in passive ownership which are plausibly uncorrelated with firm fundamentals. These allow me to causally link increases in passive ownership and decreases in pre-earnings price informativeness. This steps outside the framework of the model, where there is no such thing as an exogenous increase in passive ownership.

### **5.1 S&P 500 index additions**

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.

I obtain daily S&P 500 index constituents from Compustat. Motivated by the size and industry selection criteria, I identify a group of control firms that reasonably could have been added to the index at the same time as the treated firms. At the time of index addition, I sort firms into two-digit SIC industries. Then, within each industry, I identify the 10 firms

with the closest market capitalization to the firm that was eventually added. To be included in the final sample, control and added firms have non missing data in the two years before and after index addition. Also, the control firms must not be added to the index over the next two years. After applying all these filters, there are usually 2-3 control firms for each treated firm.

I then identify a second control group among firms that are already in the S&P 500 index. Within each 2-digit SIC industry, I identify the 10 firms with the closest market capitalization to the firm that was eventually added. I also require that these control firms do not leave the S&P 500 index over the next two years. After applying this filter, and the non-missing data filter, there is usually at least one 1 control firm for each treated firm.

In the regressions, I use index addition as an instrument for the increase in passive ownership. The first stage is:

$$\Delta Passive_{i,t} = \alpha + \beta \times Treated_{i,t} + \gamma_t + \epsilon_{i,t} \quad (34)$$

where  $\gamma_t$  is a month of index addition fixed effect. Including these fixed effects ensures I am only comparing treatment and control firms at the same point in time. The second stage is:

$$\Delta Outcome_{i,t} = \alpha + \beta \times \widehat{\Delta Passive}_{i,t} + \gamma_t + \epsilon_{i,t} \quad (35)$$

Where  $Outcome_{i,t}$  is the average pre-earnings volume, drift, or earnings day share of volatility in the two years before or after index addition. I exclude the quarter of index addition, and the quarter after index addition when computing these averages. This is to avoid index inclusion effects (see e.g. Morck and Yang (2001)), and to ensure that measures based on past averages, like abnormal volume, do not include any of the data from the pre index-addition period.

One concern is that because index addition is determined by a 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, rather than discretionary selection.

Figure 14 shows the level of passive ownership for the control firms and treated firms

around the time of index addition. Both groups of firms have similar average pre-addition changes in passive ownership, although the firms already in the index have a higher average level of passive ownership.

[Figure 14 about here.]

Table 9 contains the regression results. For comparison, I included a row with the reduced form estimates, which correspond to the equal weighted specification with all controls and fixed effects estimated in Section 4. For all three regressions, the results have the same sign as the reduced-form regressions. The estimated coefficients, however, are substantially larger than the reduced-form results. I believe the first stage 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.

[Table 9 about here.]

A natural extension is to examine firms that are 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 D.2 of the Appendix has more details on the effect of index deletion.

## 5.2 Russell 1000/2000 index rebalancing

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.<sup>35</sup> 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

---

<sup>35</sup>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.



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 et al. (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. Once 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. 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 were in the Russell 1000 for at least two years, and then switched from the Russell 1000 to the Russell 2000. To be included in the regressions, treated firms must stay in the Russell 2000 for at least two years. The control firms have June ranks between 900 and 1000 at the time the treated firms are identified. The control firms must have been in the Russell 1000 for the past two years, and must stay there for the next two years, although they are allowed to have a rank lower than 900 in the pre and post periods.

This classification involves a look-ahead bias, as I am using the ex-post changes in Russell index membership to identify changes in passive ownership. This method, however, avoids some issues discussed in Wei and Young (2017) with the most common instrumental variable approaches used in e.g. Boone and White (2015), Lin et al. (2018), Crane et al. (2016), Khan et al. (2017), Bird and Karolyi (2019) and Chen et al. (2019). In these papers, change in passive ownership is instrumented with stocks' end-of-June market capitalization, because this is *close to* what Russell uses to determine index membership. Russell actually uses end-of-May rankings to determine index membership, but does not provide these to researchers.

Wei and Young (2017) show that using June ranks, there are pre-existing differences in institutional ownership near the Russell 1000/2000 cutoff. This suggests a selection effect, rather than a treatment effect. Further, Appel et al. (2016), Crane et al. (2016) and Wei and Young (2017) discuss that using estimated May ranks leads to a weak first stage. My setup has a strong first stage: The F-statistic is over 10 (13.9), and the average increase in passive ownership associated with switching is 2.3% ( $t=11.25$ ).

Figure 15 compares the level of passive ownership around the index rebalancing date between the treated and control group. Passive ownership starts increasing in the treated firms about one quarter before they are added to the Russell 2000 index. Before this, the pre-addition changes and levels of passive ownership are similar between both groups.

[Figure 15 about here.]

For the Russell experiment, I use the same two-stage-least-squares structure as I did for the S&P 500 experiment. The only difference, is in the timing: I am comparing the two years before index reconstitution, ending in April, and the two years following reconstitution, starting in August, and then skipping the first quarter after index reconstitution. I select this window 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 prevent the trading associated with index rebalancing from influencing the regression estimates.

Table 10 contains the regression results. For comparison, I included a row with the reduced form estimates, which correspond to the equal weighted specification with all controls and fixed effects estimated in Section 4.

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 same sign, but are insignificant. Part of this could be due to the relatively short sample period (1996-2012), and the smaller number of treated and control firms. This could also be due to the increase in total volatility (i.e. the denominator of  $QVS$ ) that occurs after a firm switches from the Russell 1000 to the Russell 2000, which I discuss in the next subsection.

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.

[Table 10 about here.]

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 D.2 of the

Appendix. I show that this treatment effect is washed out by the time trend toward increased passive ownership.<sup>36</sup>

## 6 Conclusion

The goal of this paper is to not just understand *how* passive ownership affects price informativeness, but also *why*. The model reveals three competing ways that passive ownership affects which investors become informed, and what informed investors learn about. Passive ownership can make it more attractive to learn about stock-specific risks, by allowing investors to hedge their exposure to systematic risk. On the other hand, passive ownership also makes it easier to bet directly on systematic risk and makes uninformed investors better off through diversification.

The model motivates three new measures of price informativeness based on trading volume, returns and volatility around earnings announcement dates. Because of the three competing forces outlined above, the predicted effect of passive ownership on price informativeness is ambiguous. I create empirical analogues of these measures, and find that average price informativeness has declined over the past 30 years.

At the firm-level, increases in passive ownership have led to decreased 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 annual volatility on earnings days. Passive ownership is also correlated with decreased information gathering, which raises the concern of reverse causality. To rule this out, I step outside the model, and re-run the reduced-form regressions using only quasi exogenous variation in passive ownership that arises from index additions and rebalancing.

Relative to total institutional ownership, passive ownership is still relatively small, owning around 14% of the US stock market. Even at this low level, passive ownership has led to economically large changes in trading patterns, returns and the response to firm-specific news<sup>37</sup>. As passive ownership continues to grow, these changes in information and trading may be amplified, further changing the way equity markets reflect firm-specific information.

---

<sup>36</sup>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.

<sup>37</sup>See Section C.4 of the Appendix for details.

# A Appendix A: Model details, extensions and additional theoretical results

## A.1 Results based on Admati (1985)

In this subsection, I map the notation and equilibrium functions from Admati (1985) to the notation in Section 2.

Define  $Q$  as:

$$\frac{1}{\rho} \times \phi \times (S)^{-1} \quad (36)$$

where  $\phi$  is the share of rational traders who decide to become informed at cost  $c$ .

The price function is:

$$\begin{aligned} \mathbf{p} &= A_0 + A_1 \mathbf{z} - A_2 (\bar{\mathbf{x}} + \mathbf{x}) \\ A_3 &= \frac{1}{\rho} ((V)^{-1} + Q * (U)^{-1} * Q + Q) \\ A_0 &= \frac{1}{\rho} A_3^{-1} ((V)^{-1} \mu + Q(U)^{-1} \bar{\mathbf{x}}) \\ A_1 &= A_3^{-1} \left( Q + \frac{1}{\rho} Q(U)^{-1} Q \right) \\ A_2 &= A_3^{-1} \left( \mathbf{I}_n + \frac{1}{\rho} Q(U)^{-1} \right) \end{aligned} \quad (37)$$

The demand functions for informed/uninformed investors are:

$$\begin{aligned} \text{Uninformed: Demand} &= G_0 + G_{2,un} \mathbf{p} \\ \text{Informed, investor } j: \text{ Demand} &= G_0 + G_1 \mathbf{s}_j + G_{2,inf} \mathbf{p} \end{aligned} \quad (38)$$

where  $\mathbf{s}_j$  is the vector of signals received by investor  $j$  and:

$$\begin{aligned}
\gamma &= \rho (A_2^{-1} - Q) \\
G_0 &= A_2^{-1} A_0 \\
G_{2,un} &= \frac{1}{\rho} \gamma \\
G_{2,in} &= \frac{1}{\rho} (\gamma + S^{-1}) \\
G_1 &= \frac{1}{\rho} S^{-1}
\end{aligned} \tag{39}$$

The coefficients in the demand function can be used to compute investors' posterior beliefs about mean asset payoffs. For informed investors, the posterior mean conditional on signals and prices is:

$$\begin{aligned}
E_{1,j}[\mathbf{z}|\mathbf{s}_j, \mathbf{p}] &= B_{0,in} + B_{1,in}\mathbf{s}_j + B_{2,in}\mathbf{p} \\
V_{in}^a &= (V^{-1} + QU^{-1}Q + S^{-1})^{-1} \\
B_{0,in} &= \rho V_{in}^a G_0 \\
B_{1,in} &= \rho V_{in}^a G_1 \\
B_{2,in} &= \mathbf{I}_n - \rho V_{in}^a G_{2,in}'
\end{aligned} \tag{40}$$

For uninformed investors, the posterior mean conditional on prices is:

$$\begin{aligned}
E_{1,j}[\mathbf{z}|\mathbf{p}] &= B_{0,un} + B_{2,un}\mathbf{p} \\
V_{un}^a &= (V^{-1} + QU^{-1}Q)^{-1} \\
B_{0,un} &= \rho V_{un}^a G_0 \\
B_{2,un} &= \mathbf{I}_n - \rho V_{un}^a G_{2,un}'
\end{aligned} \tag{41}$$

## A.2 Supply shocks to the ETF

The supply shocks to the ETF have a different strcture than the supply shocks to the individual stocks. Define  $x_{n+1} = \tilde{x}_{n+1} + \sum_{z=1}^n x_z$  where  $\tilde{x}_{n+1}$  has the same distribution as the  $x_i$  for assets 1 to  $n$ , but is independent of  $x_i$  for all  $i$ . This implies that the supply shock

for the  $(n + 1)^{th}$  asset, the ETF, is the sum of the supply shocks to the  $n$  stocks, as well as another independent supply shock  $\tilde{x}_{n+1}$ . I define the ETF noise shocks this way based Ben-David et al. (2018) and Chinco and Fos (2019), which document transmission in noise shocks between the ETFs and the underlying assets. Assuming  $\tilde{x}_{n+1} \sim N(0, \sigma_x^2)$ , the noise shock for the  $(n + 1)^{th}$  asset has total volatility  $\sigma_{n,x}^2 = (n + 1) \times \sigma_x^2$ . The variance-covariance matrix of the noise shocks with the ETF is:  $\tilde{U} = (\Gamma')^{-1} \sigma_x^2 \mathbf{I}_{n+1} (\Gamma')^{-1}$ .

### A.3 Assumptions underlying ETF intermediary

In this sub-section, I discuss (1) why I assumed the intermediary considers the effect of her trade on expected prices and (2) why I assumed the intermediary submits a market order i.e. why her demand does not depend on prices.

The main reason for the first assumption is that I want the intermediary to be different from the informed/uninformed investors. Any of those investors could implement a trading strategy where they buy shares of the underlying stocks, and sell shares of the ETF. When risk aversion is low, informed investors will (collectively) implement a strategy like this. Given that the group of investors (informed or uninformed) ‘creating’ shares of the ETF (i.e. shorting the ETF when it is in zero average supply) is not always the same, it is not obvious how to define passive ownership. With my assumptions about the ETF creation process, passive ownership can be measured as the percent of shares of each stock purchased by the intermediary. This has the added benefit of being almost identical to the definition of passive ownership I use for the empirical section.

A way to model non-strategic ETF creation would be to have a continuum of competitive investors who can create shares of the ETF for a fixed cost (this cost maps to the creation/redemption fee charged by ETF custodians). Because these investors are competitive, in equilibrium the ETF creators will make zero economic profit, and so will be indifferent to the number of shares they create. By making the ETF creator a monopolist I get a unique solution for the size of the ETF.

The second assumption is needed because of the first assumption. At  $t = 1$ , if the intermediary could have her demand depend on prices, say through a simple linear rule, there would be an interaction between a strategic investor (the intermediary) and atomistic investors (informed and uninformed investors). On top of that, informed and uninformed

investors are learning from prices, while the intermediary, at least as she is defined now, does not. Without additional assumptions, it's not obvious that an equilibrium would exist.

## A.4 Determinants of the size of the ETF

Initially, the ETF is in zero average supply, similar to a futures contract. This means that if an investor wants to go long the ETF, there needs to be another investor taking an exactly offsetting short position in the ETF. Unlike futures contracts, however, almost all ETFs are in positive supply: few have short interest equal to 100% or more of their AUM<sup>38</sup>. The mechanism for this is that investors can take a pre-specified basket of underlying securities and give them to an ETF custodian in exchange for shares of the ETF. These shares of the ETF then trade on the secondary market.

The size of the ETF depends on the intermediary's risk aversion,  $\rho^i$ . Figure 16 shows that as the intermediary's risk aversion increases, the number of shares of the ETF decreases.

[Figure 16 about here.]

The size of the ETF also depends on  $\rho$ ,  $\sigma_n$  and the share of informed investors: if the risk-bearing capacity of the economy is low, investors will generally be willing to pay a higher price for the ETF, so the intermediary will create more shares. Figure 17 shows that as risk aversion of informed and uninformed investors increases, the equilibrium size of the ETF increases as well: The amount of the ETF created, as a function of  $\rho^i$ , shifts out to the right as we increase  $\rho$ .

[Figure 17 about here.]

## A.5 Model timeline with ETF intermediary

The model timeline for the economy with the intermediary is in Table 11. The differences from the original timeline 1 are in bold.

[Table 11 about here.]

---

<sup>38</sup>See e.g. data here on the most shorted ETFs, as of 8/1/2020 only 3 ETFs have short interest greater than or equal to 100%.

## A.6 Baseline parameters

Table 12 contains the baseline parameters. I take most of them from Kacperczyk et al. (2016) with a few exceptions: (1) I have effectively set the gross risk-free rate  $r$  to 1 because I want to de-emphasize the effect of time-discounting (2) I have 8 idiosyncratic assets, instead of 2, so investors can better attempt to replicate the systematic risk-factor with a diversified portfolio of stocks before the ETF is introduced (3) I increase the supply of the stocks. In Kacperczyk et al. (2016), the supply of the  $(n + 1)^{th}$  risk-factor i.e. the supply of the ETF in the rotated economy is 15 units, and the supply of the two stock-specific risks is 1 unit each. This implies that there is systematic risk in the economy outside the systematic risk in the stocks:  $\beta_1 \times (\text{supply of asset 1}) + \beta_2 \times (\text{supply of asset 2})$  is less than 15.

I make the total supply of all idiosyncratic assets equal to 20, and split this equally among 8 stocks. I keep the number of stocks relatively small, because if there are too many stocks, introducing the ETF has no effect. In the limit, if there were an infinite number of stocks, investors could perfectly replicate the payoff of the ETF with the underlying securities. In reality, this is stopped by trading costs, but these are absent in the model. We can view the small number of stocks as a reduced-form way of modeling transaction costs.

In this economy, increasing the share of investors who become informed (via decreasing the cost of becoming informed), decreasing the volatility of the systematic risk-factor and decreasing risk aversion have similar effects. This is because all of these changes are effectively increasing the *risk-bearing capacity* of the economy.

[Table 12 about here.]

## A.7 Sensitivity to Parameter Choice

In this sub-section, I examine how sensitive the model is to varying risk aversion and systematic risk. In Figure 18 I fix the share of investors who decide to become informed at 20% (the baseline choice in Kacperczyk et al. (2016)), and look at the effect on learning about systematic risk. As risk aversion increases, learning about systematic risk increases. This is because as risk aversion increases, the investors' diversification motive starts to dominate their profit motive. The relationship is steeper in the economy with the ETF and when the volatility of the systematic risk factor is high.



[Figure 18 about here.]

In Figure 19, I again fix the share of informed investors at 20% and vary  $\sigma_n^2$ . As expected, increasing systematic risk leads to increased learning about systematic risk. The effect is steeper when risk aversion is high and when the ETF is present.

[Figure 19 about here.]

## A.8 Numerical method for solving the model

### *Solving for $K$*

Fixing the share of informed investors, I use the following algorithm to numerically solve for the optimal  $K_i$ 's:

1. Start all investors at  $K^0$ . A simple  $K^0$  is devoting half of total attention to the systematic risk-factors, and distributing half equally among all the stock-specific risk-factors. A more sophisticated  $K^0$  is assuming the assets are independent, and solving the model using the algorithm in Kacperczyk et al. (2016)<sup>39</sup>.
2. Consider an atomistic investor  $j$  who takes  $K^0$  as given, and calculate their expected utility by deviating to  $K_j^1$  near  $K^0$ . Calculate the deviation utility for both a small increase and small decrease in the share of attention spent on the systematic risk-factor.
3. If  $j$  can be made better off, move all informed investors to  $K^1$ .
4. Iterate on steps 2 and 3 until  $j$  can no longer improve their expected utility by deviating.

At this point, it is not clear why a numerical method is needed to solve the model. Two possible alternative solution methods are (1) Adding the  $(n + 1)^{th}$  risk to Admati (1985). This will not work, as discussed in the original paper, as there is no closed form solution for prices and demands with more risks than assets. (2) Deleting the  $(n + 1)^{th}$  asset from Kacperczyk et al. (2016). This is not viable because the rotation used to isolate risk-factors and solve the model will not work if the number of risks is greater than the number of assets.

Finally, we cannot use a benevolent central planner to solve the problem: I find that in the competitive equilibrium, attention is more concentrated on a small number of risks, relative to what would maximize total expected utility for informed *and* uninformed investors.

---

<sup>39</sup>While I cannot prove uniqueness of any of these equilibria, I have not found a situation where the starting point affects the optimal attention allocation.

It also seems as though it should be possible to map the no-ETF economy to an economy with independent assets/risks via an eigendecomposition (see e.g. Veldkamp (2011)). Having done this, it would be straightforward to solve the model using the technique in Kacperczyk et al. (2016). While this is possible, it still relies on numerical methods. This is because there is no guarantee that after reversing the rotation, the solution is feasible under the proposed learning technology. See Appendix section A.12 for more details.

#### *Solving for the share of informed investors*

Because there are more risks than assets, there are no closed form solutions for  $U_{0,informed}$  and  $U_{0,uninformed}$ , but I can obtain them through simulation. Solving for  $c$  directly would be computationally intensive, as the model would have to be re-solved at each proposed combination of  $c$  and share of informed investors to check that  $U_{0,informed} = U_{0,uninformed}$ . It is easier to solve for  $c$  by creating a grid for the share of informed investors between 0 and 1. Then, at each point on the grid, compute the difference in expected utility between informed and uninformed to back out  $c$ .

#### *Solving for the size of the ETF*

I solve for the optimal  $v$  numerically using the following procedure. First, I restrict  $v$  to be greater than or equal to zero. This almost never binds, but as mentioned in the introduction of this section, almost all ETFs are in positive supply. Then, I loop over all possible values of  $v$  between 0 and  $\bar{x}$ , and select the  $v$  which maximizes the intermediary's expected utility. The expectations in Equation 15 are computed by simulating 10,000 draws of the  $z$  and  $x$  shocks for each possible choice of  $v$ .

## **A.9 Preferences: Recursive utility vs. expected utility**

In line with Kacperczyk et al. (2016), I define investors' time 0 objective function as:  $-E_0[\ln(-U_{1,j})]/\rho$  which simplifies to:  $U_0 = E_0[E_{1,j}[w_{2,j}] - 0.5\rho Var_{1,j}[w_{2,j}]]$ . This simplification comes from the fact that (1)  $w_{2,j}$  is normally distributed, and (2)  $E[\exp(ax)] = \exp(a\mu_x + \frac{1}{2}a^2\sigma_x^2)$  where  $x$  is a normally distributed random variable with mean  $\mu_x$  and standard deviation  $\sigma_x$ , and  $a$  is a constant. This objective function leads to a preference for an early resolution of uncertainty, relative to expected utility.

To see how the log transformation,  $-E_0[\ln(-U_{1,j})]/\rho$ , induces a preference for an early resolution of uncertainty relative to expected utility  $E_0[U_{1,j}]$ , I follow Veldkamp (2011) and

cast preferences as recursive utility (Epstein and Zin (1989)).

### A.9.1 Formulation as Epstein-Zin Preferences I

Start by writing down a general formulation of Epstein-Zin preferences:

$$U_t = [(1 - \beta_t)c_t^\alpha + \beta_t\mu_t(U_{t+1})^\alpha]^{1/\alpha}$$

where the elasticity of intertemporal substitution (EIS) is  $1/(1 - \alpha)$  and  $\mu_t$  is the certainty equivalent (CE) operator. I've re-labeled what is usually  $\rho$  to  $\alpha$  it to avoid confusion with the CARA risk aversion at time 1.

In my setting, all consumption happens at time 2, which simplifies things because there is no intermediate consumption. To further simplify things, set  $\beta_1 = 1$ . Choose the von Neumann-Morgenstern utility index  $u(w) = -\exp(-\rho w)$  i.e. the CARA utility at time 1. Define the certainty equivalent operator  $\mu_t(U_{t+1}) = E_t[-\ln(-U_{t+1})/\rho]$ . This  $\mu_t$  is just the inverse function of the von Neumann-Morgenstern utility index. It makes sense to call this a certainty equivalent operator because it returns the amount of dollars for sure that would yield the same utility as the risky investment. Given  $U_{1,j} = E_{1,j}[-\exp(-\rho w_{2,j})]$  and normally distributed terminal wealth,  $U_{1,j} = -\exp(-\rho E_{1,j}[w_{2,j}] + 0.5\rho^2 Var_{1,j}[w_{2,j}])$

Now, setting  $\beta_0 = 1$  and  $c_1 = 0$ :  $U_0 = [\mu_0(U_1)^\alpha]^{1/\alpha}$

Substituting in the expression for the CE operator:  $U_0 = [E_0[-\ln(-U_1)/\rho]^\alpha]^{1/\alpha}$

Putting in our expression for  $U_1$ :  $U_0 = [E_0[-\ln(\exp(-\rho E_{1,j}[w_{2,j}] + 0.5\rho^2 Var_{1,j}[w_{2,j}]))/\rho]^\alpha]^{1/\alpha}$

Simplifying:  $U_0 = [E_0[(E_{1,j}[w_{2,j}] - 0.5\rho Var_{1,j}[w_{2,j}])]^\alpha]^{1/\alpha}$

Setting  $\alpha = 1$  i.e. an infinite EIS:  $U_0 = E_0[(E_{1,j}[w_{2,j}] - 0.5\rho Var_{1,j}[w_{2,j}])]$

which matches Equation 6 in Kacperczyk et al. (2016). This shows that we can derive their utility function from Epstein-Zin preferences, but does make it totally clear what this transformation has to do with an early vs. late resolution of uncertainty.

To make things clearer, I can start with a more well-known version of Epstein-Zin preferences:  $V_t = ((1 - \beta)c_t^{1-\rho} + \beta[E_t(V_{t+1}^{1-\alpha})]^{(1-\rho)/(1-\alpha)})^{1/(1-\rho)}$

Setting  $t = 0$ ,  $c_0 = 0$ ,  $c_1 = 0$ ,  $\beta = 1$ :  $V_0 = ([E_0(V_1^{1-\alpha})]^{(1-\rho)/(1-\alpha)})^{1/(1-\rho)}$

$c^{1-\alpha}$  is a version of Constant Relative Risk Aversion (CRRA) utility. CRRA utility simplifies to log utility if relative risk aversion is equal to 1. So, with this in mind, set  $\alpha = 1$ :  $V_0 = (\exp[E_0(\ln[V_1])]^{(1-\rho)})^{1/(1-\rho)}$

Set  $\rho = 0$  (i.e. infinite EIS as above):  $V_0 = \exp[E_0(\ln[V_1])]$

This is equivalent to maximizing:  $V_0 = E_0(\ln[V_1])$  because  $\exp(x)$  is a monotone function.

In my setting:  $V_1 = E_1[\exp(-\rho w)]$  i.e. time 1 utility times -1

So the final maximization problem is:  $V_0 = -E_0(\ln[-V_1])$

With Epstein-Zin, there is a preference for an early resolution of uncertainty if  $\alpha > (1/EIS)$ . As set up here,  $\alpha = 1$  and  $1/EIS = 0$ , so investors have a preference for early resolution of uncertainty. To recover expected utility, set  $\alpha = 0$ , and then there would be no preference for early resolution of uncertainty.

#### *Why early resolution of uncertainty matters*

There are two types of uncertainty in the model: (1) uncertainty about payoffs at  $t = 2$ , conditional on signals at  $t = 1$  (2) uncertainty about portfolio you will hold at  $t = 1$  from the perspective of  $t = 0$ . With these preferences, investors are not averse to uncertainty resolved before time two i.e. are not averse to the uncertainty about which portfolio they will hold.

An intuitive way to see this is that increases in expected variance of terminal wealth,  $E_0[Var_{1,j}[w_{2,j}]]$ , linearly decrease utility. With expected utility,  $-E_0[E_1[\exp(-\rho w)]]$ , simplifies to  $-E_0[\exp(-\rho E_{1,j}[w_{2,j}] + 0.5\rho^2 Var_{1,j}[w_{2,j}])]$ . Because variance is always positive, utility is decreasing faster than linearly in expected variance.

A more nuanced argument requires a discussion of why learning about particular risks is useful. Expected excess portfolio return achieved through learning depends on the covariance between your portfolio  $q$  and asset payoffs  $f - p$ ,  $cov(q, f - p)$ . Specializing in learning about one asset leads to a high covariance between payoffs and holdings of that asset. The actual portfolio investors end up holding, however, can deviate substantially from the time 0 expected portfolio. Learning a little about every risk leads to smaller deviations between the realized and time 0 expected portfolio, but also lowers  $cov(q, f - p)$ .

With expected utility, investors are averse to time 1 portfolio uncertainty (i.e. risk that signals will lead them to take aggressive bets), so do not like portfolios that deviate substantially from  $E_0[q]$ . The utility cost of higher uncertainty from specialization offsets the utility benefit of higher portfolio returns, removing the “planning benefit” experienced by the mean-variance specification.

Recursive utility investors are not averse to risks resolved before time 2, so specialization

is a low-risk strategy. They lower their time 2 portfolio risk by loading their portfolios heavily on assets whose payoff risk will be reduced by learning.

This also shows why it is desirable to introduce a preference for an early resolution of uncertainty in endogenous learning models. Consider an investor who wants to learn about AAPL. They do this so they can hold a lot of Apple (AAPL) when it does well, and hold little AAPL when it does poorly. An expected utility investor would be hesitant to learn too much about AAPL, because the fact that their portfolio will vary substantially depending on the signal they get seems risky to them.

## A.10 Extensions

### A.10.1 Extension 1: Endogenous capacity choice

In the main body of the paper, the exogenous learning margin is a binary choice: Pay the fixed cost  $c$  and become informed, or stay uninformed. This can be made into a continuous choice as follows: Fix the share of informed investors, but allow them to optimally choose their total attention  $K$ . I consider two functional forms for the cost of adding capacity: (1) Linear:  $c(K) = aK + b$  and (2) Convex  $c(K) = aK^2 + b$ .

The effect of varying  $K$  depends on the share of informed agents. Figure 20 shows two features of this extended version of the model when  $\rho = 0.25$  and  $\sigma_n = 0.25$ : (1) For any share of informed investors, as you increase total attention, investors devote less attention to systematic risk (2) For any amount of total attention, as you increase the share of informed investors, they devote less attention to systematic risk.

These patterns arise because in economies with medium to low *risk-bearing capacity*, investors follow a threshold rule for learning. When the total amount of information in the economy is small, either because capacity is low, or because the share of informed investors is low, investors devote all their attention to systematic risk. This is the market-timing channel at work: when investors are risk averse, they care more about systematic risk than idiosyncratic risk, because idiosyncratic risk can be diversified away.

Eventually, the price of the ETF becomes informative enough that investors want to start spreading out their attention. Given that  $\sigma = 0.55 > \sigma_f = 0.25$ , there is more money to be made betting on individual stocks than on the ETF. So once the total information in the economy is large enough, informed investors want to learn more about stock specific risks.

[Figure 20 about here.]

To numerically solve this version of the model, I loop over values of  $K$ , and find the point where the ex-ante utility of the informed and uninformed investors is equal, given  $c(K)$ .

I find the predictions of this extensions for the extensive learning margin and all three measures of price informativeness unchanged from the baseline model. If the risk-bearing capacity of the economy is low, increasing passive ownership leads investors to choose less capacity, and allocate that capacity mostly to systematic risk. This is true for both the linear and convex  $c(K)$ .

### A.10.2 Extension 2: Heterogeneous assets

In the baseline version of the model, I assume all informed investors have the same  $K_{i,j} = K_i$ . In addition, I assume that assets 1 to  $n$  have the same: (1) Mean (2) Systematic risk (3) Idiosyncratic risk (4) Supply shock variance. These assumptions reduce an otherwise  $n$  dimensional problem (the  $(n + 1)^{th}$  dimension is accounted for by the total information constraint) to a two dimensional problem: Informed investors must only decide to allocate  $K_{n+1}$  attention to systematic risk, and  $(1 - K_{n+1})/n$  to each idiosyncratic risk-factor.

Suppose now that each asset  $i$  now has the payoff:

$$z_i = a_i + \beta_i f + \eta_i \quad (42)$$

where  $\beta_i$  and  $var(\eta_i)$  is different for each asset. In this setting, informed investors' choice is not just a trade-off between learning about systematic and idiosyncratic risk. To solve for information choice in this version of the model, I need to modify the numerical method:

1. Start all investors at  $K^0$
2. Consider an atomistic investor  $j$  who takes  $K^0$  as given, and considers their expected utility by deviating to  $K_j^1$  on a  $n \times n$  dimensional grid around  $K^0$ . Even though there are  $(n + 1)$  risks to learn about, we don't need the  $(n + 1)^{th}$  dimension because of the total information constraint.
3. Calculate the gradient numerically at  $K^0$  using this grid of expected deviation utilities. Then, move  $j$  on the grid in the direction of the gradient.
4. If  $j$ 's expected utility increased, move all informed investors to  $K_j^1$

5. Iterate on steps 2-4 until  $j$  can no longer improve their expected utility by deviating.

When the ETF is present, this method is able to match closed form solutions from Kacperczyk et al. (2016) with heterogeneous  $\beta_i$ 's. For  $n > 3$ , however, this method can take a long time to find the solution. Allowing for heterogeneous assets does not drastically change the model's predictions for the effect of passive ownership on pre-earnings volume, pre-earnings drift or earnings-day volatility, so I focus on the symmetric asset case in the main body of the paper.

## A.11 Additional theoretical results

### A.11.1 Effect of introducing the ETF on investors' posterior mean and variance

Introducing the ETF changes the way investors form beliefs about asset payoffs. Define  $\mathbf{s}_p = \mathbf{z} + \epsilon_p$  as the signal about asset payoffs contained in prices. From the price function, this can be written as:  $\mathbf{s}_p = A_1^{-1}(\mathbf{p} - A_0)$ , which implies that  $\epsilon_p = A_1^{-1}A_2(\bar{\mathbf{x}} + \mathbf{x})$  and  $\Sigma_p = A_1^{-1}A_2U$  where  $U$  is the variance-covariance matrix of supply shocks. This implies that  $\mathbf{s}_p \sim N(0, \Sigma_p)$ . Without the ETF:

$$\underbrace{\widehat{\Sigma_j^{-1}}}_{\text{Posterior Precision}} = \underbrace{V^{-1}}_{\text{Prior Precision}} + \underbrace{\Sigma_p^{-1}}_{\text{Price Precision}} + \underbrace{S_j^{-1}}_{\text{Signal Precision}} \quad (43)$$

With the ETF, investors observe  $s_{p,n+1}$  i.e. the signal about payoff of the  $(n+1)^{th}$  asset contained in asset prices. This will change  $\Sigma_p^{-1}$  i.e. the price precision, but nothing else. This is because fixing attention allocation, introducing the ETF has no effect on  $S_j^{-1}$  for assets 1 to  $n$ . For any asset  $i$ ,  $s_{i,j} = (f + \epsilon_{f,j}) + (\eta_i + \epsilon_{i,j})$ , so  $var(s_{i,j}) = var(\epsilon_{f,j} + \epsilon_{i,j}) = var(\epsilon_{f,j}) + var(\epsilon_{i,j})$  by independence of the signal noises.

When the ETF is not present, the posterior mean of  $f$  will be:

$$\underbrace{E_{1,j}[\mathbf{z}]}_{\text{Posterior Mean}} = \underbrace{\hat{\Sigma}_j}_{\text{Posterior Variance}} \times \left( \underbrace{S_j^{-1}}_{\text{Precision on Asset Signals}} \mathbf{s}_j + \underbrace{\Sigma_p^{-1}}_{\text{Price Precision}} \mathbf{s}_p \right) \quad (44)$$

With the ETF, investors can separately weigh their signal for  $f$  by its own precision:

$$\underbrace{E_{1,j}[\mathbf{z}]}_{\text{Posterior Mean}} = \underbrace{\hat{\Sigma}_j}_{\text{Posterior Variance}} \times \left( \underbrace{S_j^{-1}}_{\text{Precision on Risk-Factor Signals}} \mathbf{s}_j + \underbrace{\Sigma_p^{-1}}_{\text{Price Precision}} \mathbf{s}_p \right) \quad (45)$$

where the terms that have changed are in color. To see how this works, I apply the eigendecomposition in Veldkamp (2011) to isolate the risk-factors. Pre-multiplying  $\mathbf{z}$  by  $\Gamma$ , creates synthetic assets exposed to only one risk-factor each:

$$\begin{aligned} \mathbf{z} = \mu + \Gamma\eta &\leftrightarrow \tilde{\mathbf{z}} = \Gamma^{-1}\mu + \eta \\ \tilde{z}_i &= \eta_i + \tilde{\epsilon}_i \text{ for } i = 1, \dots, n \end{aligned} \quad (46)$$

With this rotation, the supply of the synthetic assets is  $(\Gamma')^{-1}(\bar{\mathbf{x}} + \mathbf{x})$ , but at this point, the signals may still be correlated. After another transformation to make the signals independent, I can solve for the equilibrium in this economy using the numerical technique in Kacperczyk et al. (2016)<sup>40</sup>, and then rotate back to the economy with payoffs  $\mathbf{z}$  and signals  $\mathbf{s}$ . In this rotated economy, it is clear that investors are going to independently use the  $(n+1)^{th}$  signal, and the price of the  $(n+1)^{th}$  asset to learn about  $\mathbf{z}$ , something they cannot do in the no-ETF world.

To quantify the effect of introducing the ETF on investors' posterior precisions, Table 13 contains selected entries of  $\hat{\Sigma}$ . Introducing the ETF always increases the precision of both the informed and uninformed for assets 1 to  $n$ .

[Table 13 about here.]

### A.11.2 Effect of passive ownership on risk premia

Fixing the share of investors who become informed in equilibrium, introducing the ETF almost always decreases expected returns in the economy. This is not surprising, as the ETF increases the information in the economy: it adds an  $(n+1)^{th}$  public signal, the price of the ETF. Table 14 shows that introducing the ETF decreases average asset returns, as long as risk aversion and the volatility of systematic risk are not too high. Once we allow the share of

---

<sup>40</sup>I would like to thank the authors for sharing their solution code with me.



informed investors to vary, however, risk premia can actually increase. This is because as the number of informed investors in the economy decreases, the effective risk-bearing capacity of the economy decreases, so risk premia must increase.

I view the effect of the ETF on risk premia as more of a modeling artifact than a testable prediction, and want to take out this effect when studying price informativeness. To do this, I work with market-adjusted returns: I calculate the returns of each asset as the actual return, minus the market returns, as discussed in Campbell et al. (2001). Market-adjusted returns are also used for all the empirical exercises in this paper. Whether or not the ETF is present, the market is defined as the average return of all the stocks, to ensure an apples-to-apples comparison<sup>41</sup>.

[Table 14 about here.]

### **A.11.3 Expected utility of informed and uninformed investors**

Table 15 contains information on the percentage difference in expected utility between informed and uninformed investors when the ETF is and is not present.

[Table 15 about here.]

### **A.11.4 Sensitivity of demand to prices**

As shown in Section 2, when investors get good signals about a particular asset, they invest more in it. At the same time, they hedge this bet by either (1) shorting an equal-weighted portfolio of all the other stocks when the ETF is not present (2) shorting the same number of shares of the ETF when it is present.

Similar to the hedging demand from informed investors' private signals, all investors use prices as a signal, and thus may do a similar hedging. Table 16 fixes the cost of becoming informed. Table 17 fixes the share of investors becoming informed.

[Table 16 about here.]

[Table 17 about here.]

---

<sup>41</sup>The results are unaffected the market is defined as the return of the ETF when it is present.

### A.11.5 Effect of varying baseline learning $\alpha$

One of the effects of setting  $\alpha$  to larger values than the baseline of 0.001 in Section 2, is that a kink forms in the relationship between the cost of becoming informed and the share of investors who decide to learn when the ETF is present. To the right of the kink, the cost of becoming informed is high, so relatively few investors are becoming informed. Given that systematic risk affects all assets, informed investors initially devote all their attention to learning about this risk-factor.

To the left of the kink, learning about the systematic risk-factor has become crowded, and informed investors start devoting some attention to the individual-asset risks. All informed investors get some information for *free* about each risk-factor from  $\alpha$ . This means that there is a meaningful difference between devoting zero attention to a risk-factor, and devoting a small positive amount of attention to that same risk-factor.

Figure 21 focuses on the case where  $\rho = 0.25$  and  $\sigma_n^2 = 0.2$ . The top panel shows two things: (1) The relationship between the cost of becoming informed, and the share of attention devoted to systematic risk [blue dots]. To the right of the kink, all attention is being devoted to the systematic risk-factor. (2)  $U_{1,j}$  i.e. the time one objective function for informed [red squares] and uninformed investors [green triangles]. One of the counter-intuitive features of the kink is that the line is *steeper* once investors are devoting some attention to the idiosyncratic assets. For both informed and uninformed investors, the lines become steeper to the left of the kink.

The second panel shows why the slope changes: To the right of the kink informed and uninformed investors are making roughly the same profits on stocks, but informed investors are making significantly larger profits on the ETF. To the left of the kink, informed investors gain an advantage over uninformed investors on the individual stocks. This increases the relative benefit of becoming informed, which can explain the changes in slopes around the kink.

[Figure 21 about here.]

## A.12 Representation as economy with independent assets

Consider an alternative economy with no ETF, where all asset payoffs are:

$$z_i = a_i + \eta_i \quad (47)$$

i.e. with no systematic component, but instead of having the  $\eta_i$  be *i.i.d.*, have them correlated in a way that replicates the structure of the payoffs. This model can be solved the same way as the model in Section 2.

There is no guarantee, however, that there will be an apples-to-apples learning comparison with the economy when the ETF is present. This happens when the solution to the rotated model proposes values for  $K_i$  which do not satisfy the total information constraint. For example, suppose we have two assets and three risks. Using the notation in the appendix of Kacperczyk et al. (2016):

Define:  $\Sigma^{1/2}$  = Square root of  $V$ , the variance-covariance matrix of payoffs

Define:  $\Sigma_s = S$ , the variance-covariance matrix of signals

$$\text{Define: } \Sigma_s^1 = \Sigma^{-1/2} \times \Sigma_s \times \Sigma^{-1/2} \quad (48)$$

We can re-write:  $\Sigma_s = \Sigma^{1/2} \times G \times L \times G \times \Sigma^{1/2}$

where  $G$  and  $L$  come from the eigen-decomposition of  $\Sigma_s^1$

Define orthogonal signal matrix:  $\tilde{\Sigma}_s = G' \times \Sigma^{-1/2} \times \Sigma_s \times (\Sigma^{-1/2})' \times G$

This implies that:

$$\tilde{\Sigma}_s = \begin{bmatrix} 1/(\alpha + \tilde{K}_1) & 0 \\ 0 & 1/(\alpha + \tilde{K}_2) \end{bmatrix} \quad (49)$$

After solving the model, the optimal  $\tilde{K}_i$  rotated back to the original economy may require  $K_i$  that do not satisfy  $\sum_i \tilde{K}_i \leq K$ , where  $K$  is the original total information constraint.

In the next subsection, I outline how to ensure the learning technologies are comparable between the economy with and without the ETF.

### A.12.1 Solving a rotated version of the model

This rotated version of the model can be solved numerically in the following steps:

1. Guess an initial total attention  $K$  for informed investors
2. Solve the orthogonal model with this total attention constraint
3. Loop over possible attention choices in the un-rotated model
4. See if optimal attention from the rotated model matches the guess after rotation i.e.  $\Sigma_e = GL^*G'$  where  $GLG' = \Sigma_e$  is the eigen-decomposition of the signal precision matrix and  $L^*$  is the optimal precision matrix in the rotated model
5. Loop over all possible total attention allocations for the orthogonal model until it matches desired total attention in the un-rotated model

If assets are not independent, the following additional transformation to  $\Sigma_e$  is needed in step 4:  $\Sigma_e = \Sigma^{1/2}GL^*G\Sigma^{1/2}$ , where  $\Sigma$  is the covariance matrix of asset payoffs.

## A.13 Effects of Introducing the ETF vs. Effect of Growing Passive Ownership

Introducing the ETF, even in zero average supply, completes the market. As shown in other parts of this Appendix, introducing the ETF has effects on investors' demand functions, posterior beliefs, and expected utility. In this subsection, I outline an empirical exercise which better maps to the *introduction* of ETFs, rather than the growth of passive ownership.

### A.13.1 Introduction of sector ETFs

Sector ETFs are ETFs that track specific industries, rather than the market as a whole. While there are many sector ETFs, the most well known are State Street's Sector SPDR Funds. Table 18 contains a list of the all the sector SPDRs. These funds were introduced in waves: The first set was introduced in 1999. The second wave, which were all sub-sects of the S&P 500 were introduced in 2005 and 2006. The third wave, also subsets of the S&P 500 was introduced in 2011, while the final few were introduced in 2015 and later. As of June 2020 there is over \$170 Billion invested in these products.

[Table 18 about here.]

The introduction of sector ETFs captures some of the features of introducing the ETF in the model. These are low fee products (expense ratios less than 50 basis points), so they

make it easier to trade on systematic risk. These are also heavily used by hedge funds to short/hedge sector risks. According to Goldman Sachs Hedge Fund Monitor (2016), hedge funds are net short XRT, XLY, XBI, XOP, XLI, XLF, XLV, XLU and XLE – and this is only among sector ETFs they explicitly listed in their report. Further, these net short positions are not due to small long positions. For example, in XLF (the Financial Select Sector SPDR), hedge funds have net \$712 million long and \$2.4 *billion* short. The large short interest in many ETFs by hedge funds may be due to the ETFs relatively low borrowing cost. According to Deutsche Bank, shorting SPY (the largest S&P 500 ETF) usually costs about 40 basis points, while shorting something riskier like the US consumer staples sector ETF (XLP) can run up to 72 basis points.

I examine the effect of the introduction of the original set of sector ETFs in 1999 on price informativeness. There are three groups of firms to compare (1) firms which were in the ETFs (2) firms in the same sector as the ETF, but were not part of the ETF basket (3) firms in sectors without ETFs.

The firms which were added to the sector ETFs are mostly firms in the largest 20% of each industry. To construct a better control group, I split firms up into quintiles of market capitalization by industry. The two groups of treated firms are (1) those in the ETF (2) those in the same 3-digit SIC industry as firms in the ETF and in the top 20% of market capitalization for these industries, but not in the ETF. The control group is going to be firms in 3-digit SIC industries that do not have sector ETFs, but are still in the top 20% of market capitalization for their own industry.

The empirical estimates I am trying to match are from the following regression:

$$Outcome_{i,t} = \alpha + \beta \times Treated_{i,t} \times Post_t + \gamma_t + \epsilon_{i,t} \quad (50)$$

where *Outcome* is pre-earnings volume, pre-earnings drift, or share of volatility on earnings days.  $Post_t = 1$  for all year/quarters after the first quarter of 1999. I omit the second quarter of 1999 for the volume/drift regressions (which use quarterly data) in case of a temporary liquidity shock to these stocks as the result of the ETFs being introduced. For the volatility regression (which uses annual data), I omit all of 1999 for the same reason. *Treated* equals 1 if a firm was in one of the ETFs, or in a sector with an ETF. It is equal to zero otherwise. The regression also includes time fixed effects,  $\gamma_t$ , and because of these time fixed effects,

there is no uninteracted  $Post_t$ . Observations are weighted by lagged market capitalization. Standard errors are clustered at the firm-level.

Table 19 contains the regression results. After the introduction of sector ETFs, the treated firms had a decrease in pre-earnings volume, a decrease in pre-earnings drift, and an increase in the share of annual volatility on earnings days. The effect is slightly stronger among the treated firms that were members of the new sector ETFs<sup>42</sup>. This is consistent with the sector ETFs decreasing stock-level price informativeness. A calibration is in the row below the regression results. The calibrated parameters are: (1) informed share before ETF introduction at 90%, after ETF introduction 70% (2) Risk aversion  $\rho$  at 0.15, and (3) volatility of systematic risk  $\sigma_n^2$  at 0.3. In the calibration, the ETF is in zero average supply, designed to capture the fact that these ETFs were small when they were first introduced.

[Table 19 about here.]

The calibration is able to quantitatively match the changes in volume and volatility, but can only qualitatively match the results for the pre-earnings drift. Part of this is due to the fact that in the data, I am using returns over 22 trading days to construct the drift, while in the model, there is only 1 day before the earnings announcement. The concept of information being slowly incorporated into prices might be better suited to a Kyle (1985)-style model, than the model in this paper.

[Figure 22 about here.]

## B Appendix B: Robustness of stylized facts

### B.1 Data details

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.

---

<sup>42</sup>See Figure 22 for a visual version of these regressions.

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 when ruling out the influence of Regulation Fair Disclosure 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%.

## B.2 Alternative definitions of pre-earnings volume

Rather than look at the 22 days before an earnings announcement, I expand the analysis to 60 trading days before the earnings announcement. 60 trading days roughly corresponds to the time between earnings announcements. A concern with the regression specification in Equation 24 (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 23 shows the cross-sectional median pre-earnings volume, which exhibits the same decline in pre-earnings volume as Figure 9.

[Figure 23 about here.]

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 *average* 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, either in aggregate, or for stocks with high passive ownership.

### B.3 Alternative definitions of the pre-earnings drift

One alternative way to define a pre-earnings drift measure is using squared returns, rather than signed returns. Define the drift volatility,  $DV$ , as the ratio of the squared earnings day return to the squared return on all days since last earnings announcement plus squared earnings day return. This is similar to  $QVS$ , but is defined for each individual earnings announcement, rather than a whole year.

If  $DV$  is near one, almost all volatility occurs on earnings days, while if it is near zero, almost all volatility happens between earnings days. In a Kyle (1985)-type model,  $DV$  would decrease with the precision of the insider's signal<sup>43</sup>. All my baseline pre-earnings drift results are robust to switching any drift magnitude measure for  $DV$ .

### B.4 Decomposition of earnings-days' share of annual volatility

Figure 24 decomposes the rise of  $QVS$  into rise in the numerator (volatility on earnings days) and the denominator (volatility on all days). The trend in  $QVS$  was driven by a simultaneous increase in the numerator, and decrease in the denominator.

[Figure 24 about here.]

### B.5 Value weights vs. equal weights

[Figure 25 about here.]

[Figure 26 about here.]

[Figure 27 about here.]

---

<sup>43</sup>I thank Alex Chinco for making his two-period Kyle code available on his website.



## B.6 Placebo tests for stylized facts

In this subsection, I conduct two placebo tests for the stylized test (1) Using days between earnings announcements as placebo announcement dates (2) Using FOMC meetings as placebo announcement dates.

### B.6.1 Days between earnings announcements

This section replicates Figures 9 (decrease in pre-earnings volume), 11 (decrease in pre-earnings drift) and 13 (increase in earnings day volatility), except replaces the true earnings dates with placebo earnings dates 22 days before the actual announcements. In all three cases, there is no trend toward decreased informativeness on the placebo earnings dates.

[Figure 28 about here.]

[Figure 29 about here.]

### B.6.2 Systematic Information Announcement Days

The other set of placebo announcement days I focus on are FOMC announcement dates. I obtain FOMC announcement dates from Gorodnichenko and Weber (2016). To create an apples-to-apples comparison with the anticipated nature of earnings announcements, I restrict the sample to scheduled FOMC meetings. I compute versions of pre-earnings volume, and pre-earnings drift for these FOMC dates. The only difference is that I use a  $\pm 10$  day around each announcement, instead of  $\pm 22$  days, to avoid overlap as there are 8 scheduled meetings per year. Share of volatility on FOMC meeting dates is the sum of squared returns on those dates, divided by the sum of squared returns on all dates.

Figure 30 shows the trends in volume around FOMC announcement dates. There is no drop before the announcement in the last third of the sample. Figure 31 shows the pre-FOMC announcement drift. There is no upward/downward trend throughout the sample. Figure 32 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 global financial crisis.

[Figure 30 about here.]

[Figure 31 about here.]

[Figure 32 about here.]

[Figure 33 about here.]

## **C Appendix C: Robustness of reduced-form results**

### **C.1 Placebo tests for reduced-form regressions**

[Table 20 about here.]

### **C.2 Alternative explanations for the decline of price informativeness**

In this subsection, I discuss three threats to identification in my main baseline regressions (1) Regulation Fair Disclosure (2) The rise of algorithmic trading (3) Omitted variables.

#### **C.2.1 Regulation Fair Disclosure (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 9, 12 and 13, 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 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.

Another possibility is that Reg FD changed the way insiders (directors or senior officers) behaved, or led to changes in the enforcement of insider trading laws. Time-series changes in enforcement should be accounted for by year fixed-effects. To rule out the insider behavior channel, I used the Thompson Insiders data to compute insider buys/sells as a percent of total

shares outstanding for each firm in my dataset. Insider buys and sells have been decreasing since the mid-1990's i.e. before Reg FD. Both average annual buys and sells went down slightly more for stocks with increases in passive ownership, but this effect is only weakly statistically significant. I then examined insider buys/sells in 22 day windows before/after earnings announcements. Both buys and sells have decreased before and after earnings announcements, broadly following the trend toward decreased insider activity. There is no statistically significant relationship between increases in passive ownership and changes in insider buys/sells before or after earnings announcements.

Finally, 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, I re-run the reduced-form regressions using only post-2000 data in Tables 21, 22 and 23. 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.

[Table 21 about here.]

[Table 22 about here.]

[Table 23 about here.]

### **C.2.2 The Rise of algorithmic trading (AT) activity**

Weller (2017) shows that Algorithmic Trading (AT) activity is negatively correlated with pre-earnings price informativeness. His proposed mechanism is algorithmic traders back-run informed traders, reducing 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 algorithmic traders in the trends toward decreased pre-earnings price informativeness as I cannot directly observe AT activity, and only have reasonable AT activity measures between 2012-2018. 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, however, should not affect my results, as I condition on firm size in all the reduced-form regressions (2) High ETF ownership will attract algorithmic traders 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 2018. 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) Truncate 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 almost entirely the result of the year restrictions.

I re-run all the reduced form regressions, but restrict to the matched sample with MIDAS, and include the 4 measures of HFT activity. As a sanity check, I first re-run the baseline 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”.

Tables 24, 25 and 26 contain the regressions with AT controls. All the results in this matched sub-sample are qualitatively unchanged from Tables 4, 5 and 6. For the pre-earnings drift, and earnings day share of volatility, adding the AT activity controls does not significantly change the coefficient on change in passive ownership. For volume, the value-weighted results are robust to including the AT controls, while the equal weighted result is the right sign, but insignificant. This could imply (1) The equal-weighted volume

result has become weaker over time, as the coefficient in column 2 of Table 24 is 1/3 the size of the same coefficient in column 2 of Table 4. Given that I can only include the AT controls in the matched subsample, it is hard to fully disentangle this effect (2) Part of the AT activity measure is mechanically correlated with passive ownership because, for example, ETFs attract algorithmic traders implementing ETF arbitrage. I show AT activity and passive ownership are positively correlated in Table 27 (described in detail below) (3) 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.

[Table 24 about here.]

[Table 25 about here.]

[Table 26 about here.]

One possible reason for the drop in statistical significance when including the AT activity measures is a strong correlation between passive ownership and AT activity. To test this, I calculate an AT activity score as the first principal component of the 4 AT measures in Weller (2017). Table 27 runs a regression of the AT activity score on the level of and changes in passive. Across almost all specifications, the relationship is positive and statistically significant. In unreported results, I find AT activity also increases in stocks after they are added to the S&P 500.

[Table 27 about here.]

### C.3 Possible omitted variables

In addition to identification concerns, the reduced-form regressions could suffer from omitted variable bias. Most of 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<sup>44</sup>, 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.

Another possibly omitted variable is the quantity of ETF rebalancing. The changes in pre-earnings volume could be driven by mechanical changes in systematic trading rules by ETFs, as in Chinco and Fos (2019). I find this unlikely, as ETFs typically rebalance on a calendar frequency, not around particular firms' earnings announcements, which may be scattered throughout a calendar quarter. In unreported results, I find that the drop in volume before earnings announcements is robust to including the ETF rebalancing or imbalance measures of Chinco and Fos (2019) on the right-hand-side of the reduced-form regression.

It is also possible that firms with high passive ownership anticipate increased volatility on earnings days, or amplified responses to earnings news (as I show in Section C.4 of the Appendix) and as a result, release earnings information when the market is closed. To test this, I form three groups of earnings announcements (1) Weekday, before 4PM EST (2) Weekday, after 4PM EST (3) Weekend or trading holiday. I then run a multinomial logistic regression of these categories on passive ownership. Passive ownership does not have significant predictive power in this regression, once I control for time trends, and differences in firm characteristics correlated with passive ownership like market capitalization.

A final omitted variable is the composition and importance of systematic and idiosyncratic information in the economy. The growth of passive management could be a response to a secular increase in the importance of systematic risks, and increased demand for exposure to these factors. If this were true, it might be rational to focus on learning about factor risk, rather than firm-specific risk, which could explain my results. I find this unlikely for several reasons.

Figure 34 replicates the volatility decomposition in Campbell et al. (2001), extending the

---

<sup>44</sup>There was a sharp size cutoff before the rule change in 2006, see e.g. Wei and Young (2017).

analysis to 2017. Between 1990 and 2017, there does not appear to have been a significant increase in the market or industry components of total risk. It is still possible, however, that this does not fully account for changes in systematic risk, if those changes occur on a small number of days, like FOMC announcements, as the figure is composed of slow-moving averages. Figure 35 re-constructs Figure 1, but using FOMC announcement dates instead of earnings announcement dates. Between the 1990's and 2010's there was no major change in returns or trading volume around these systematic information release dates at the stock level, also inconsistent with an increase in importance of systematic risk.

[Figure 34 about here.]

[Figure 35 about here.]

## C.4 Additional Result 1: Stock responses to earnings surprises

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 their 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.

### C.4.1 Trends in earnings responses

To measure trends in earnings responses, I run two types of regressions. The baseline comes from Kothari and Sloan (1992):

$$r_{i,t} = \alpha + \beta \times SUE_{i,t} + controls + \epsilon_{i,t} \quad (51)$$

I also design an earnings-response regression which allows for asymmetry between positive and negative surprises:

$$r_{i,t} = \alpha + \beta_1 \times SUE_{i,t} \times \mathbf{1}_{SUE_{i,t} > 0} + \quad (52)$$

$$\beta_2 \times |SUE_{i,t}| \times \mathbf{1}_{SUE_{i,t} < 0} + controls + \epsilon_{i,t} \quad (53)$$

The results of running these regressions year-by-year, and taking a 5-year moving average of the  $\beta$ 's are in Figure 36. Over the past 30 years, earnings responses have increased.

[Figure 36 about here.]

#### C.4.2 Effect of passive ownership on earnings responses

To test the predictions in Buffa et al. (2014), I run the following regression:

$$r_{i,t} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 (SUE_{i,t} \times Passive_{i,t}) + \xi X_{i,t} + \text{Fixed Effects} + e_{i,t} \quad (54)$$

Here,  $r_{i,t}$  denotes the market-adjusted return on the effective quarterly earnings date.  $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$ .<sup>45</sup> Controls in  $X_{i,t}$  include 1-year lagged passive ownership, market capitalization, growth of market capitalization from  $t-1$  to  $t$ , idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. Fixed effects include year/quarter and firm.

I also run a version of Equation 54, breaking SUE into two variables: They are equal to SUE if SUE is positive (negative), and zero otherwise. I then interact these variables with passive ownership, to account for a possibly asymmetric effect of passive ownership on positive and negative news.

Table 28 contains the regression results. Consistent with the Buffa et al. (2014) model, firms with a high share of passive ownership are more responsive to earnings news, especially if that news is negative<sup>46</sup>.

[Table 28 about here.]

---

<sup>45</sup>Results are similar when calculating SUE relative to IBES estimates using the method in Anson et al. (2012).

<sup>46</sup>An alternative explanation is that passive ownership, in particular ETFs, has reduced short sale constraints, as in Palia and Sokolinski (2019). In unreported results, I calculate the shorting of individual stocks through ETFs, and I find this cannot explain the larger response of stocks with high passive ownership to negative earnings news.



## C.5 Additional result 2: Price of options that span earnings announcements

Given the increase in volatility on earnings days, especially for stocks with high passive ownership, it is natural to believe that ex-ante measures of uncertainty, like options prices, should also reflect this change. To quantify the change in option prices around earnings announcements, I follow the method in Kelly et al. (2016). For each earnings announcement  $\tau$ , I select 3 expiration dates,  $a$ ,  $b$ , and  $c$ , where  $a$  the last expiration before the earnings announcement date,  $b$  is the first expiration after the announcement and  $c$  is expiration after  $b$ . For each of these dates, I average implied volatility (IV) across all options with  $|\Delta| \in (0.4, 0.5)$  in a 20-day window starting 1 day before  $\tau$ . I also choose corresponding 20-day windows before  $a$  and  $c$  to match time to expiration.

This measure was originally designed for S&P 500 options, so I have to modify it to be comparable across firms:

$$IVD_\tau = \frac{\overline{IV}_b}{0.5 (\overline{IV}_a + \overline{IV}_c)} \quad (55)$$

I also calculated the *Variance Risk Premium* and *Slope* measures from Kelly et al. (2016), but given that I am using individual equity options, I had the longest/most reliable sample using the *IVD* measure.

I constructed a second measure of ex-ante uncertainty based on Dubinsky et al. (2006). Suppose there is one predictable announcement before an option expires:

$$\text{Before event IV} = \sigma^2 + \frac{1}{T_i} (\sigma_j^\mathbb{Q})^2 \quad (56)$$

I estimate earnings uncertainty using two options with different maturities,  $T_1 < T_2$ , on the day before earnings are released:

$$(\sigma_{term}^\mathbb{Q})^2 = \frac{\sigma_{t,T_1}^2 - \sigma_{t,T_2}^2}{T_1^{-1} - T_2^{-1}} \quad (57)$$

For each firm/earnings announcement, I average this measure over the 3 strikes closest to the money. To make this comparable across firms, I divide by average IV:  $\frac{(\sigma_{term}^\mathbb{Q})^2}{\overline{IV}_b}$ :

To construct these option-based measures of earnings uncertainty, I used daily implied

volatility data from OptionMetrics. To make sure I am working with reliable data, I restrict to S&P 500 firms, discard all options with  $\frac{ask}{bid} > 5$ , or zero open interest and filter for firms which have at least 15 years of non-missing options data.

### C.5.1 Trends

Figure 37 shows that both measures of option prices around earnings announcements increased substantially between the 1990's and 2010's.

[Figure 37 about here.]

### C.5.2 Effect of passive ownership on options that span earnings announcements

I run the following regression to see if changes in passive ownership can explain the increases in *IVD* and *Term*:

$$\Delta_{(t,t-5)}Outcome_{i,t} = \alpha + \beta\Delta_{(t,t-5)}Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \quad (58)$$

Controls in  $X_{i,t}$  include institutional ownership, lagged institutional ownership, market capitalization, lagged market capitalization. Fixed effects include industry, year and firm. The results are in Table 29. Although the effects are not always statistically significant, increases in passive ownership are correlated with increases in both the *IVD* and *Term* measures of earnings uncertainty.

[Table 29 about here.]

## D Appendix D: Robustness of quasi-experimental results

### D.1 Effect of passive ownership vs. effect of index inclusion

One concern is that the quasi-experimental results are driven by index inclusion effects (see e.g. Elliott et al. (2006)), rather than the increase in passive ownership associated with index inclusion. For example, the general consensus is that when a firm is added to an index,

its stock returns become more correlated with the stock returns of other firms in that index. If this changed the distribution of stock returns on non-earnings announcement dates, it could bias my results for the pre-earnings drift, and share of annual volatility on earnings dates.

To rule out this channel, I examine the effect of index additions/rebalancing on CAPM beta, CAPM R-squared, idiosyncratic volatility and total volatility. For the S&P 500 experiment, the results are in Figure 38. I find that when a firm is added to the S&P 500 index, its CAPM beta increases (marginally statistically significant) and its CAPM R-squared increases (statistically significant). There is no effect on the magnitude of CAPM residuals (i.e. idiosyncratic volatility), but there is an increase in total volatility (statistically significant). This last fact is consistent with Ben-David et al. (2018) and Chinc0 and Fos (2019), where being a member of an ETF basket leads to additional volatility.

For the Russell experiment, the results are in Figure 39. Firms which switch have a significant increase in CAPM beta, and a marginally significant decrease in CAPM R-squared. Switching firms also have significant increases in idiosyncratic volatility and total volatility. Firms which switch from the Russell 1000 to the 2000 are shrinking, so these last two facts are consistent with bad news usually being associated with increased volatility.

These facts imply that index inclusion effects are likely working against my results on the share of volatility on earnings days. If total volatility increases after index addition/rebalancing, the denominator of  $QVS$  should increase, and shrink  $QVS$ . This may explain why the volatility results are insignificant for the Russell experiment: In the post period, treated firms have an increase in total volatility that was two times as large as the increase in total volatility for the corresponding firms in the S&P 500 experiment.

[Figure 38 about here.]

[Figure 39 about here.]

## D.2 Alternative quasi-exogenous changes in passive ownership

### D.2.1 S&P 500 index deletions

In Section 5, 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 exogeneity assumption is likely violated, 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.<sup>47</sup>

Figure 40 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 is that stocks on the margin are still relatively large. Passive ownership increased for all stocks over my sample, especially the larger ones. The weak and temporary treatment effect suggests that index deletion would be a weak instrument for change in passive ownership.

[Figure 40 about here.]

### **D.2.2 Moving from the Russell 2000 to the Russell 1000**

Similar to 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

---

<sup>47</sup>All the 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.

still satisfies the exogeneity assumption, as moving from firm 1001 to 999 may have nothing to do with firm fundamentals.

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 15 shows the problem with this setup: the treatment is small and temporary. The common pattern between moving from the Russell 2000 to the 1000, and 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 index rebalancing.

[Figure 41 about here.]

### D.2.3 Blackrock's acquisition of Barclays Global Investors

Another well-known source of quasi-exogenous variation in passive ownership is Blackrock's acquisition of Barclays' iShares ETF business in December 2009. This is not an ideal setting for testing my hypothesis because: (1) My theory has no predictions for the effects of increased concentration of ownership among passive investors (Azar et al. (2018), Massa et al. (2018)) (2) While there may have been a *relative* increases in flows to iShares ETFs, relative to all other ETFs (Zou (2018)), I do not find a significant increase in overall ETF ownership for the stocks owned by iShares funds. Given that my right-hand-side variable of interest is the percent of shares owned by passive investors, the model has no predictions for the effect of moving dollars from iShares ETFs to non-iShares ETFs.

## References

- Admati, A. R. (1985). A noisy rational expectations equilibrium for multi-asset securities markets. *Econometrica: Journal of the Econometric Society*, pages 629–657.
- Admati, A. R. and Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, 1(1):3–40.
- Akerlof, G. A. (1978). The market for lemons: Quality uncertainty and the market mechanism. In *Uncertainty in economics*, pages 235–251. Elsevier.
- Anson, M. J., Chambers, D. R., Black, K. H., Kazemi, H., Association, C., et al. (2012). *CAIA level I: an introduction to core topics in alternative investments*, volume 709. John Wiley & Sons.
- Appel, I. R., Gormley, T. A., and Keim, D. B. (2016). Passive investors, not passive owners. *Journal of Financial Economics*, 121(1):111–141.
- Azar, J., Schmalz, M. C., and Tecu, I. (2018). Anticompetitive effects of common ownership. *The Journal of Finance*, 73(4):1513–1565.
- Bai, J., Philippon, T., and Savov, A. (2016). Have financial markets become more informative? *Journal of Financial Economics*, 122(3):625–654.
- Balchunas, E. (2016). *The Institutional ETF Toolbox: How Institutions Can Understand and Utilize the Fast-Growing World of ETFs*. John Wiley & Sons.
- Ball, R. and Brown, P. (1968). An empirical evaluation of accounting income numbers. *Journal of accounting research*, pages 159–178.

- Ben-David, I., Franzoni, F., and Moussawi, R. (2018). Do etfs increase volatility? *The Journal of Finance*, 73(6):2471–2535.
- Bhattacharya, A. and O’Hara, M. (2018). Can etfs increase market fragility? effect of information linkages in etf markets. *Effect of Information Linkages in ETF Markets (April 17, 2018)*.
- Bird, A. and Karolyi, S. A. (2019). Retraction: Governance and taxes: Evidence from regression discontinuity. *The Accounting Review*.
- Boone, A. L. and White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, 117(3):508–533.
- Buffa, A. M., Vayanos, D., and Woolley, P. (2014). Asset management contracts and equilibrium prices. Technical report, National Bureau of Economic Research.
- Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have individual stocks become more volatile? an empirical exploration of idiosyncratic risk. *The Journal of Finance*, 56(1):1–43.
- Chen, S., Huang, Y., Li, N., and Shevlin, T. (2019). How does quasi-indexer ownership affect corporate tax planning? *Journal of Accounting and Economics*, 67(2-3):278–296.
- Chinco, A. and Fos, V. (2019). The sound of many funds rebalancing.
- Coffee Jr, J. C. (2007). Law and the market: The impact of enforcement. *U. pa. L. rev.*, 156:229.
- Cong, L. W. and Xu, D. (2016). Rise of factor investing: asset prices, informational efficiency, and security design. In *29th Australasian Finance and Banking Conference*.

- Crane, A. D., Michenaud, S., and Weston, J. P. (2016). The effect of institutional ownership on payout policy: Evidence from index thresholds. *The Review of Financial Studies*, 29(6):1377–1408.
- Dávila, E. and Parlatore, C. (2019). Trading costs and informational efficiency. Technical report, National Bureau of Economic Research.
- Dubinsky, A., Johannes, M., et al. (2006). Fundamental uncertainty, earning announcements and equity options. In *Working Paper*.
- Elliott, W. B., Van Ness, B. F., Walker, M. D., and Warr, R. S. (2006). What drives the s&p 500 inclusion effect? an analytical survey. *Financial Management*, 35(4):31–48.
- Epstein, L. G. and Zin, S. (1989). Risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. *Econometrica*, 57(4):937.
- Fama, E. F., Fisher, L., Jensen, M. C., and Roll, R. (1969). The adjustment of stock prices to new information. *International economic review*, 10(1):1–21.
- Farboodi, M. and Veldkamp, L. (2017). Long run growth of financial technology. Technical report, National Bureau of Economic Research.
- Foster, G., Olsen, C., and Shevlin, T. (1984). Earnings releases, anomalies, and the behavior of security returns. *Accounting Review*, pages 574–603.
- Garleanu, N. and Pedersen, L. H. (2018). Active and passive investing. *Available at SSRN 3253537*.
- Glosten, L. R., Nallareddy, S., and Zou, Y. (2016). Etf trading and informational efficiency of underlying securities.

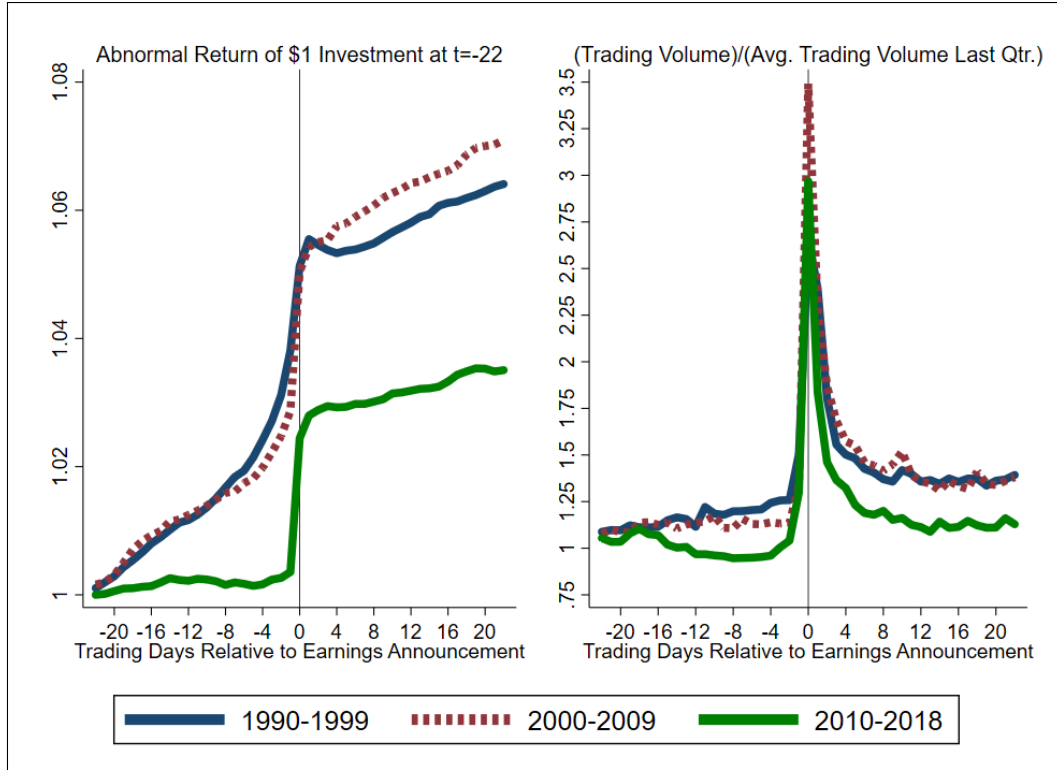


- Gorodnichenko, Y. and Weber, M. (2016). Are sticky prices costly? evidence from the stock market. *American Economic Review*, 106(1):165–99.
- Grossman, S. J. and Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American economic review*, 70(3):393–408.
- Israeli, D., Lee, C. M., and Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? an information perspective. *Review of Accounting Studies*, 22(3):1048–1083.
- Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A rational theory of mutual funds’ attention allocation. *Econometrica*, 84(2):571–626.
- Kacperczyk, M. T., Nosal, J. B., and Sundaresan, S. (2018). Market power and price informativeness. *Available at SSRN 3137803*.
- Kelly, B., Pástor, L., and Veronesi, P. (2016). The price of political uncertainty: Theory and evidence from the option market. *The Journal of Finance*, 71(5):2417–2480.
- Khan, M., Srinivasan, S., and Tan, L. (2017). Institutional ownership and corporate tax avoidance: New evidence. *The Accounting Review*, 92(2):101–122.
- Kothari, S. P. and Sloan, R. G. (1992). Information in prices about future earnings: Implications for earnings response coefficients. *Journal of Accounting and Economics*, 15(2-3):143–171.
- Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica: Journal of the Econometric Society*, pages 1315–1335.
- Lin, Y., Mao, Y., and Wang, Z. (2018). Institutional ownership, peer pressure, and voluntary disclosures. *The Accounting Review*, 93(4):283–308.

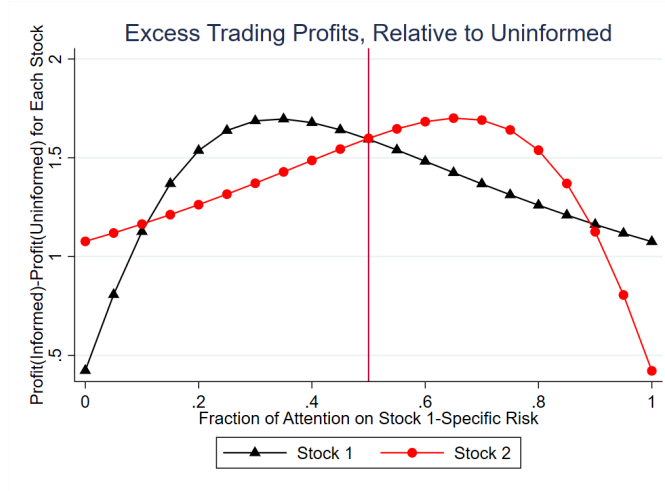
- Loughran, T. and McDonald, B. (2017). The use of edgar filings by investors. *Journal of Behavioral Finance*, 18(2):231–248.
- Malikov, G. (2018). Information, participation, and passive investing. *Participation, and Passive Investing (December 15, 2018)*.
- Massa, M., Schumacher, D., and Wang, Y. (2018). Who is afraid of blackrock?
- Mauboussin, M., Callahan, D., and Majd, D. (2017). Looking for easy games. Technical report, Credit Suisse.
- McLean, R. D. and Pontiff, J. (2016). Does academic research destroy stock return predictability? *The Journal of Finance*, 71(1):5–32.
- Merton, R. C. (1987). A simple model of capital market equilibrium with incomplete information.
- Milgrom, P. and Stokey, N. (1982). Information, trade and common knowledge. *Journal of economic theory*, 26(1):17–27.
- Morck, R. and Yang, F. (2001). The mysterious growing value of s&p 500 membership. Technical report, National Bureau of Economic Research.
- Palia, D. and Sokolinski, S. (2019). Passive asset management, securities lending and stock prices. *Securities Lending and Stock Prices (February 15, 2019)*.
- Schipper, K. (1989). Earnings management. *Accounting horizons*, 3(4):91.
- S&P, D. J. I. (2017). S&p u.s. indices methodology. Technical report.

- Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*. Princeton University Press.
- Wang, J. (1994). A model of competitive stock trading volume. *Journal of political Economy*, 102(1):127–168.
- Wei, W. and Young, A. (2017). Selection bias or treatment effect? a re-examination of russell 1000/2000 index reconstitution.
- Weller, B. M. (2017). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies*, page hhx137.
- Zou, Y. (2018). Lost in the rising tide: Etf flows and valuation.

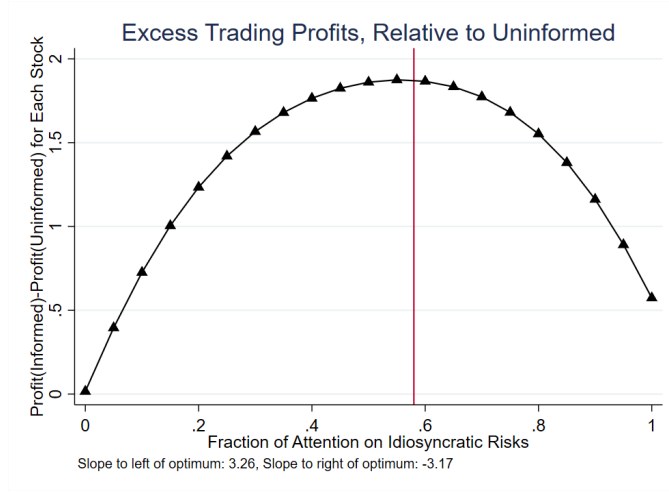
## List of Figures



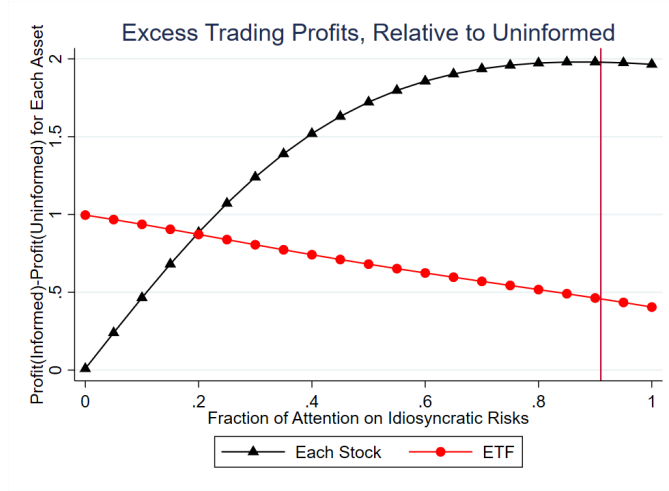
**Figure 1. Returns and Trading Volume around Earnings Announcements.** Each plot represents the cross-sectional average for firms in the top decile of Standardized Unexpected Earnings (SUE). SUE is defined as:  $SUE_{i,t} = \frac{Earnings_{i,t} - Earnings_{i,t-4}}{\sigma_{(t-1,t-8)}(Earnings_{i,t} - Earnings_{i,t-4})}$ . Deciles of SUE are calculated each quarter. Abnormal returns are returns minus the returns on the CRSP value-weighted index. Abnormal volume is volume divided by the firm-level average volume over the past quarter.



**Figure 2. Two Stock Example, No Systematic Risk.** Vertical red line denotes optimal attention allocation. All other points are not equilibrium outcomes. 20% of investors are informed. Residual attention is on Stock 2-specific risk.  $\rho = 0.1$ ,  $\sigma^2 = 0.55$

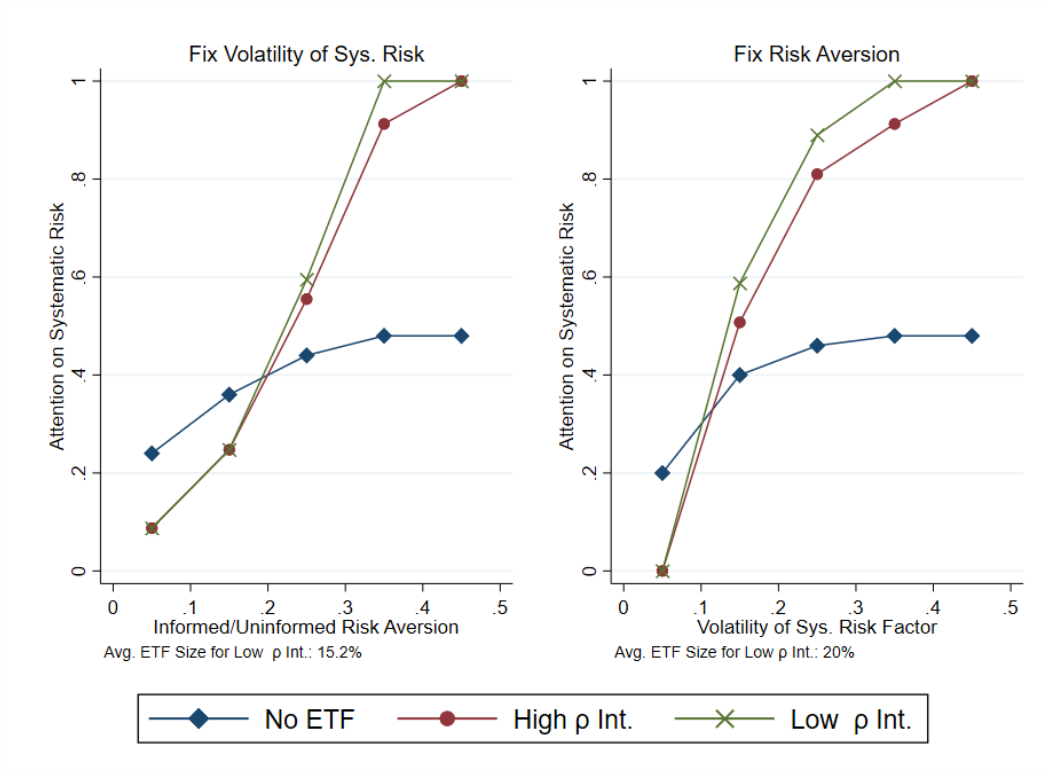


**Figure 3. .Two stock example, systematic risk, no ETF** Vertical red line denotes optimal attention allocation. All other points are not equilibrium outcomes. 20% of investors are informed. Attention on stock-specific risks is equal. Residual attention is on systematic risk-factor.  $\rho = 0.1$ ,  $\sigma_f^2 = 0.2$ ,  $\sigma^2 = 0.55$

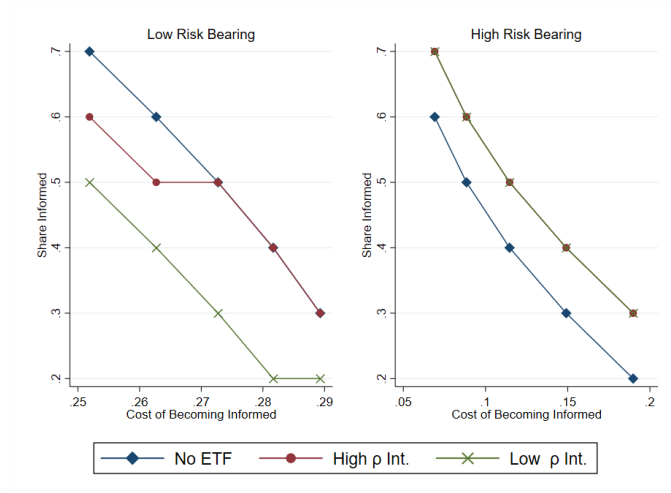


**Figure 4. Two stock example, systematic risk, ETF present.** Vertical red line denotes optimal attention allocation. All other points are not equilibrium outcomes. 20% of investors are informed. Residual attention is on systematic risk-factor. ETF is in zero average supply.  $\rho = 0.1$ ,  $\sigma_f^2 = 0.2$ ,  $\sigma^2 = 0.55$

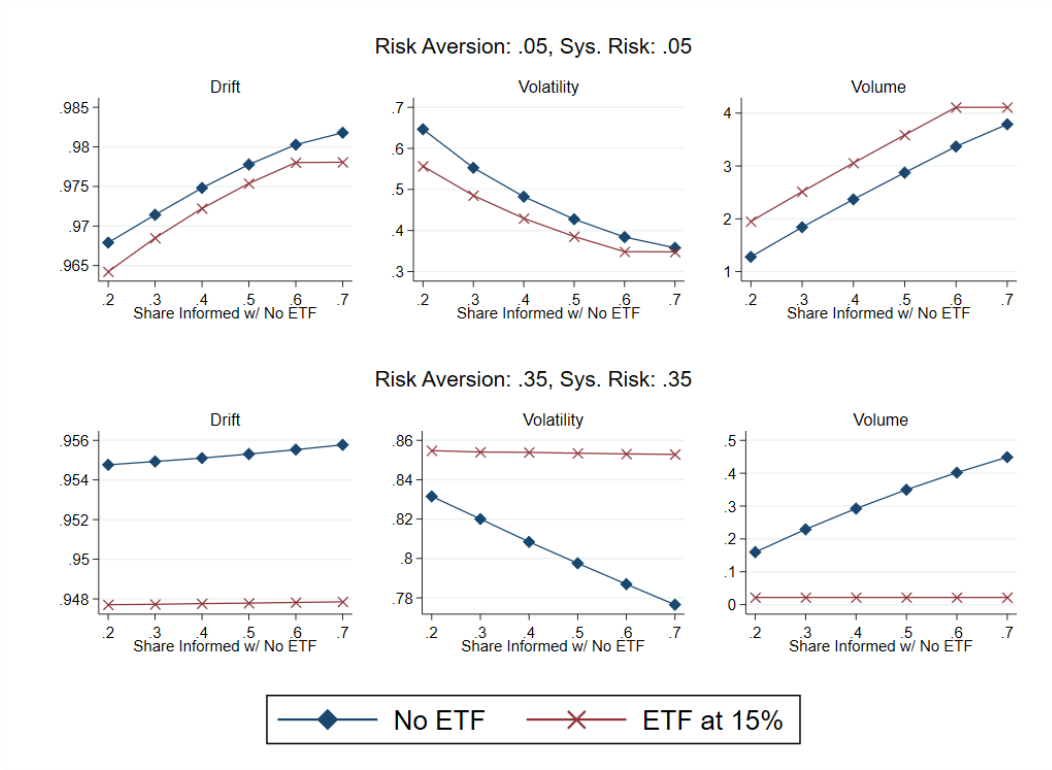




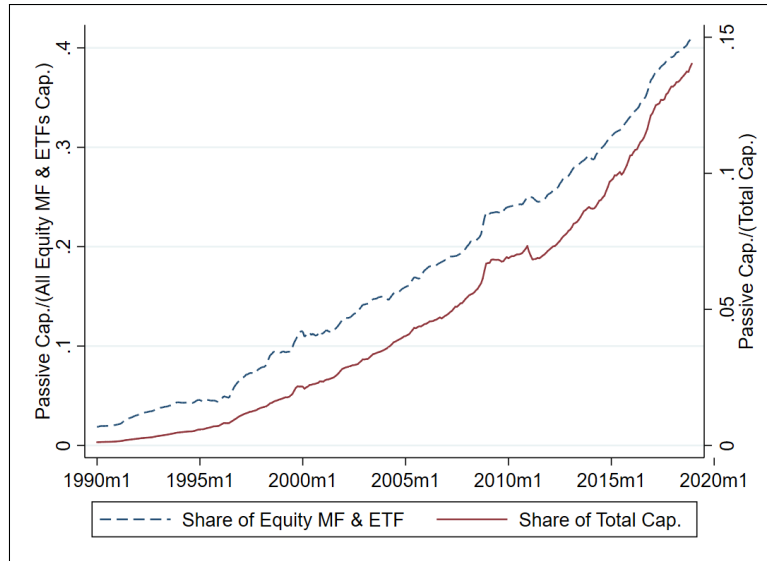
**Figure 5. Effect of Introducing the ETF on Attention Allocation (fixed share informed).** In both panels, the share of informed agents is fixed at 60%. In the left panel,  $\sigma_f = 0.35$ . In the right panel,  $\rho = 0.35$ . The y-axis reports the share of investors' attention devoted to the systematic risk-factor.



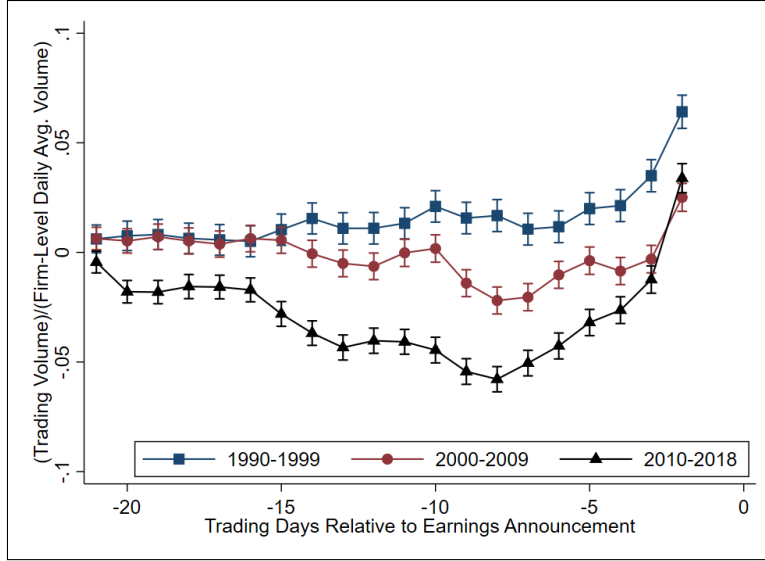
**Figure 6. Effect of passive ownership on extensive learning margin.** Left panel:  $\sigma_f = 0.25$ ,  $\rho = 0.15$ . Right panel:  $\sigma_f = 0.05$ ,  $\rho = 0.05$ . The x-axis is the cost in dollars of becoming informed. The y-axis reports the share of investors who become informed in equilibrium at this cost.



**Figure 7. Effect of Passive Ownership on Price Informativeness.** Top panel:  $\rho = 0.05$  and  $\sigma_n = 0.05$ . Bottom panel:  $\rho = 0.35$  and  $\sigma_n = 0.35$ . The cost of being informed is set so to match the share of agents who become informed when the ETF is not present.



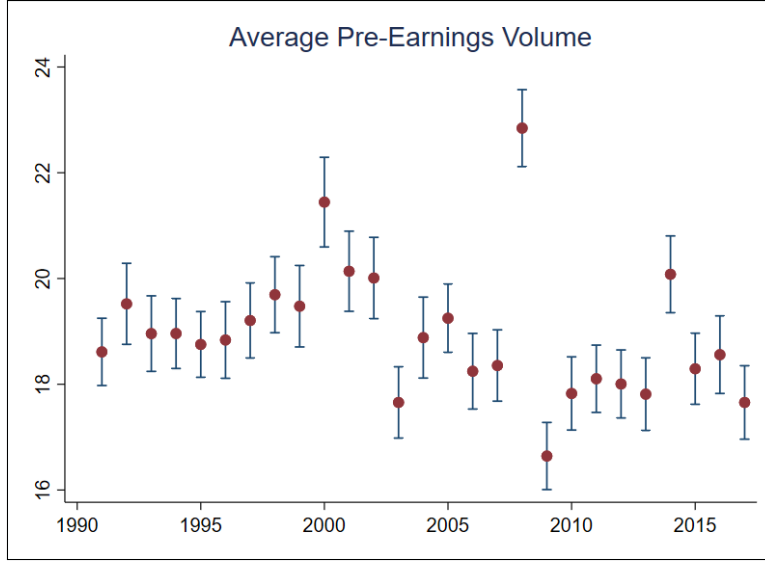
**Figure 8. The Rise of Passive Ownership: 1990-2018.** Passive ownership is defined as all index funds, all ETFs, and all mutual funds with “index” in the name. Index funds are identified using the index fund flag in the CRSP mutual fund data. Total equity mutual fund and ETF assets is the sum of all stock holdings in the Thompson S12 data that can be matched to CRSP. Total market capitalization includes all ordinary common shares traded on major exchanges in CRSP.



**Figure 9. Decline of Pre-Earnings Volume.** Plot of  $\beta_k$  estimated from the regression:

$$AV_{i,t+\tau} = \alpha + \sum_{k=-21}^{22} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau}$$

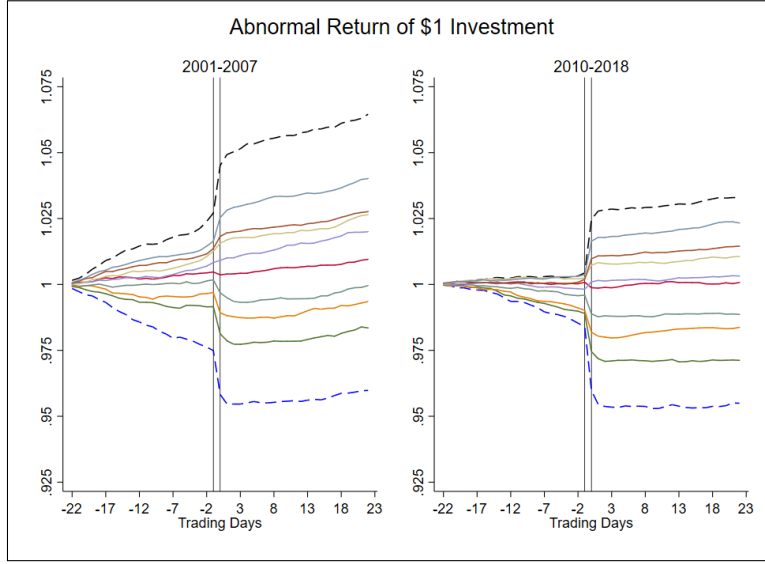
Bars represent a 95% confidence interval around the point estimate. Standard errors are clustered at the firm level. Observations are weighted by 1-month lagged market capitalization.



**Figure 10. Average Cumulative Abnormal Pre-Earnings Volume by Year.** This figure plots coefficients from a regression of  $CAV_{i,t}$  on a set of year dummy variables where

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

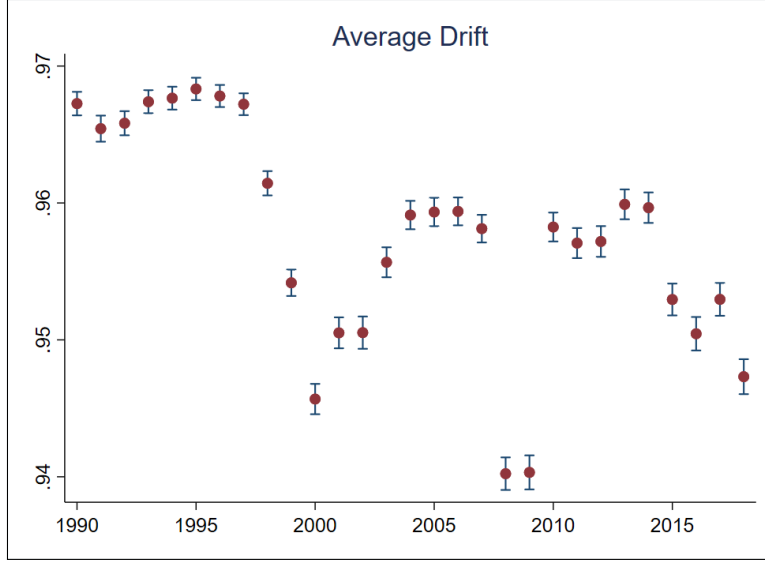
the sum of abnormal trading volume from  $t - 22$  to  $t - 1$  for firm  $i$  around earnings date  $t$ . Observations are weighted by 1-month lagged market capitalization. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.



**Figure 11. Decline of Pre-Earnings Drift by SUE Decile.** Each quarter, I sort firms into deciles on:

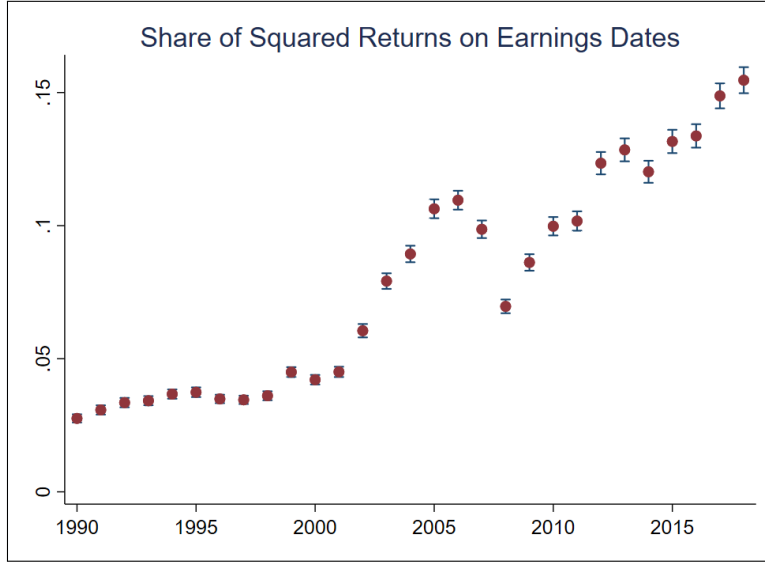
$$SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$$

Each line represents the cross-sectional average market-adjusted return of \$1 invested at  $t=-22$ . 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. The solid lines represent the averages for deciles 2 to 9.

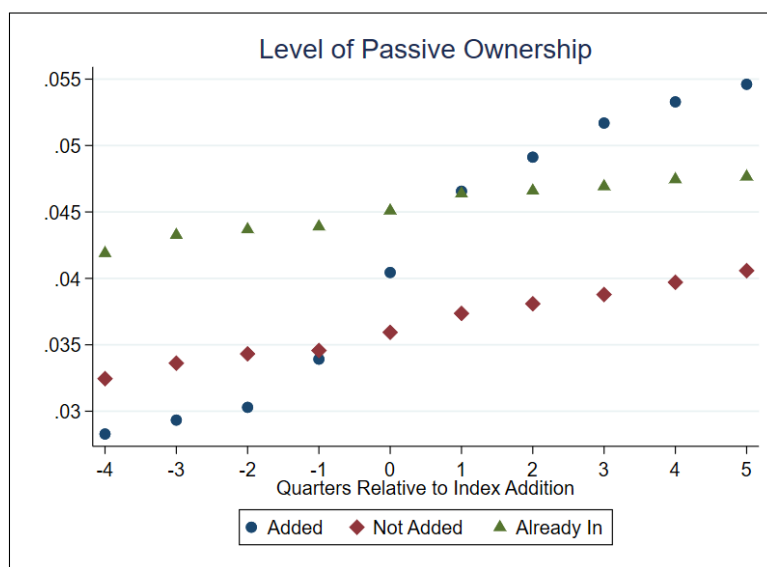


**Figure 12. Decline of Average Pre-Earnings Drift.** This figure plots coefficients from a regression of pre-earnings drift on a set of year dummy variables where  $DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases}$ . A value near 1 implies most earnings information is incorporated in prices before the announcement date, while lower values denote less informative pre-earnings announcement prices. Observations are weighted by 1-month lagged market capitalization. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.

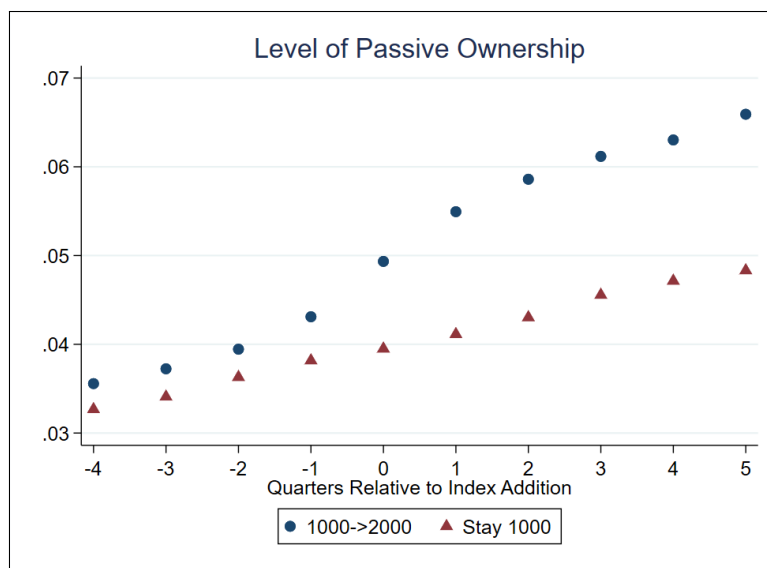




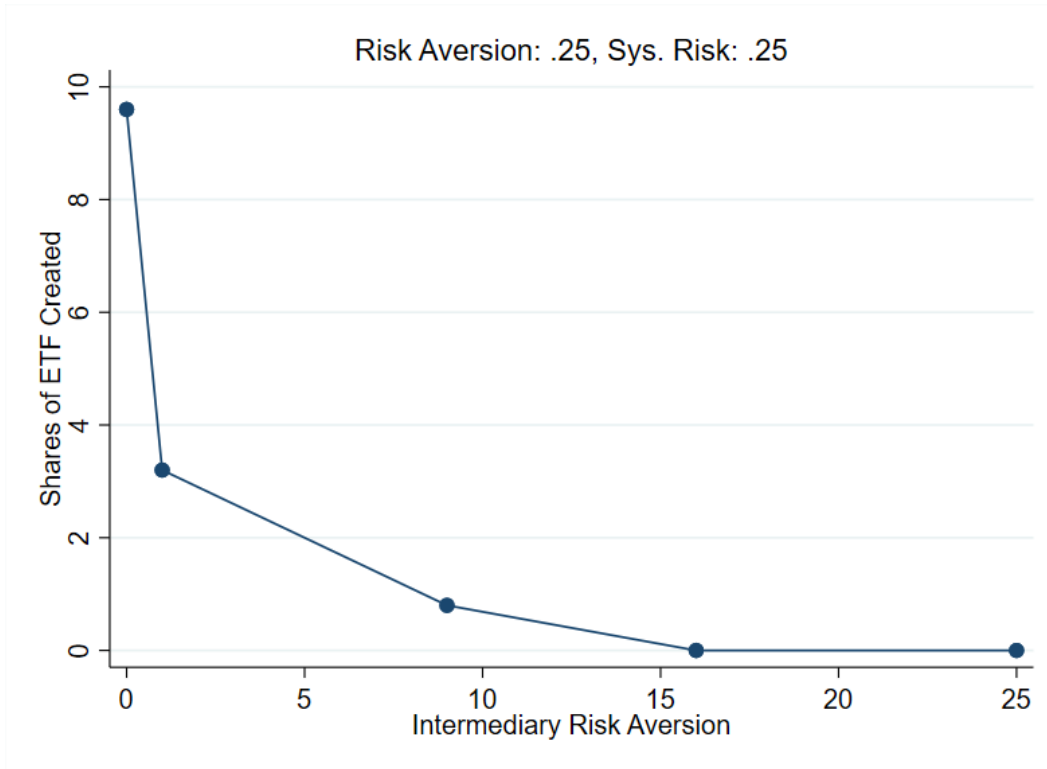
**Figure 13. Increase in Earnings Day Volatility.** This figure plots coefficients from a regression of QVS on a set of year dummy variables. For firm  $i$  in year  $t$  the quadratic variation share (QVS) is defined as:  $QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , where  $r$  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  $t$ . Observations are weighted by 1-year lagged market capitalization. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.



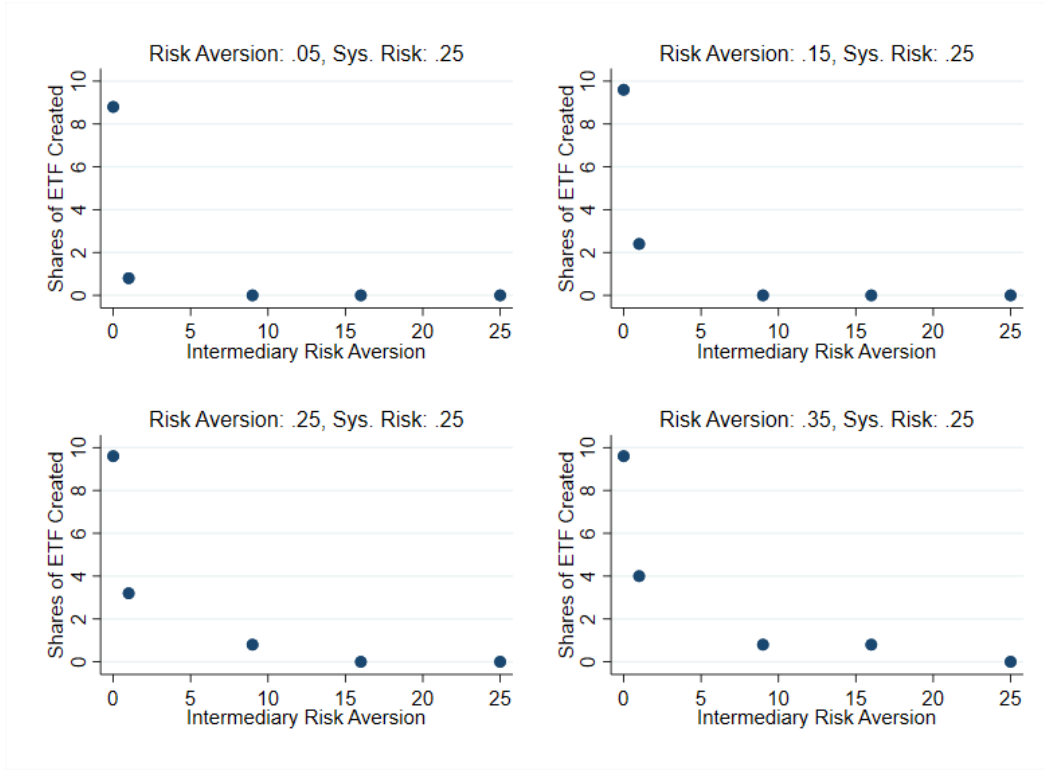
**Figure 14. S&P 500 Index Addition and Changes in Passive Ownership.** Average level of passive ownership for control firms out of the index (“Not Added”), control firms in the index (“Already In”) and treated firms (“Added”).



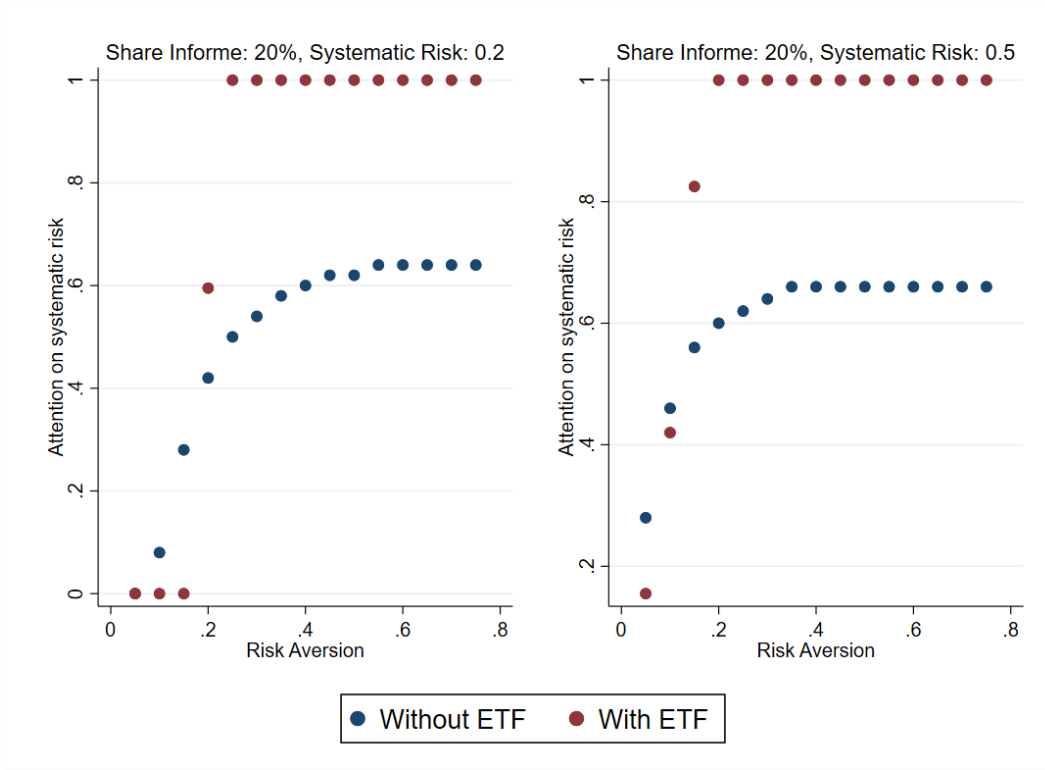
**Figure 15. Russell 1000/2000 Reconstitution and Changes in Passive Ownership.** Average level and increase in passive ownership for control firms and firms moved from the Russell 1000 to the Russell 2000. Control firms are all firms in the Russell 3000 ranked 900 to 1100 that did not move from the 1000 to the 2000 or from the 2000 to the 1000.



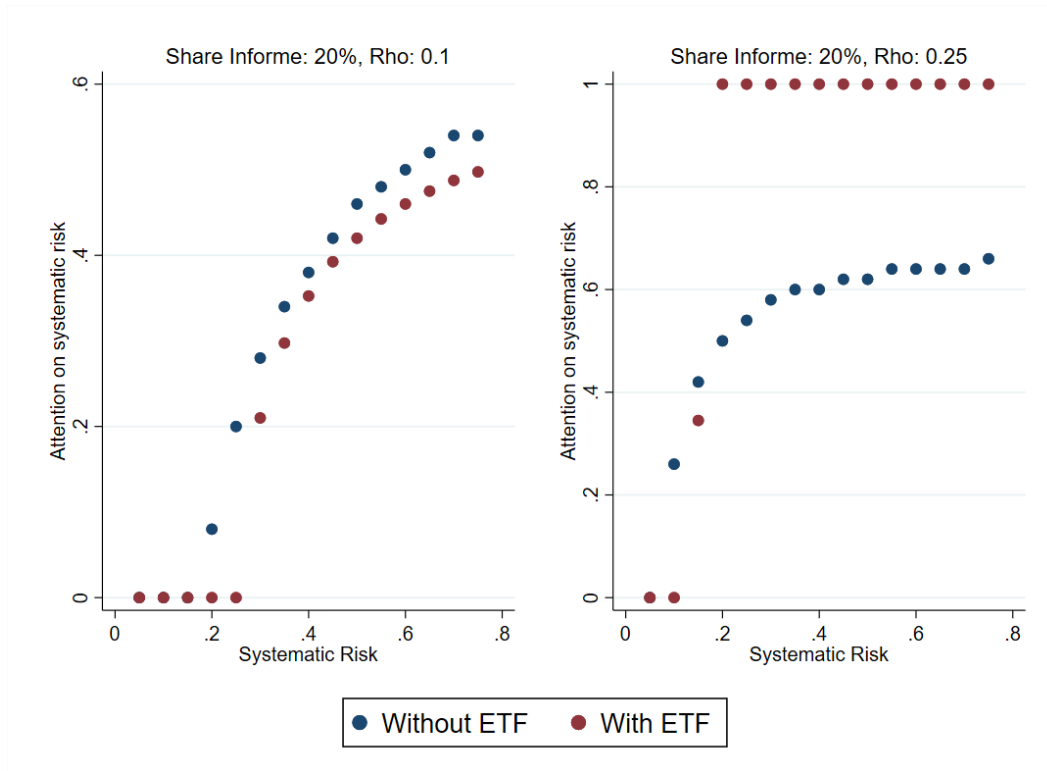
**Figure 16. Relationship between size of the ETF and the intermediary's risk aversion.** Risk aversion of informed/uninformed investors  $\rho = 0.25$ . Volatility of the systematic risk factor  $\sigma_f = 0.25$ . The share of informed investors is set to 50%. If the ETF owned all the shares of the underlying stocks, it would have 20 shares outstanding.



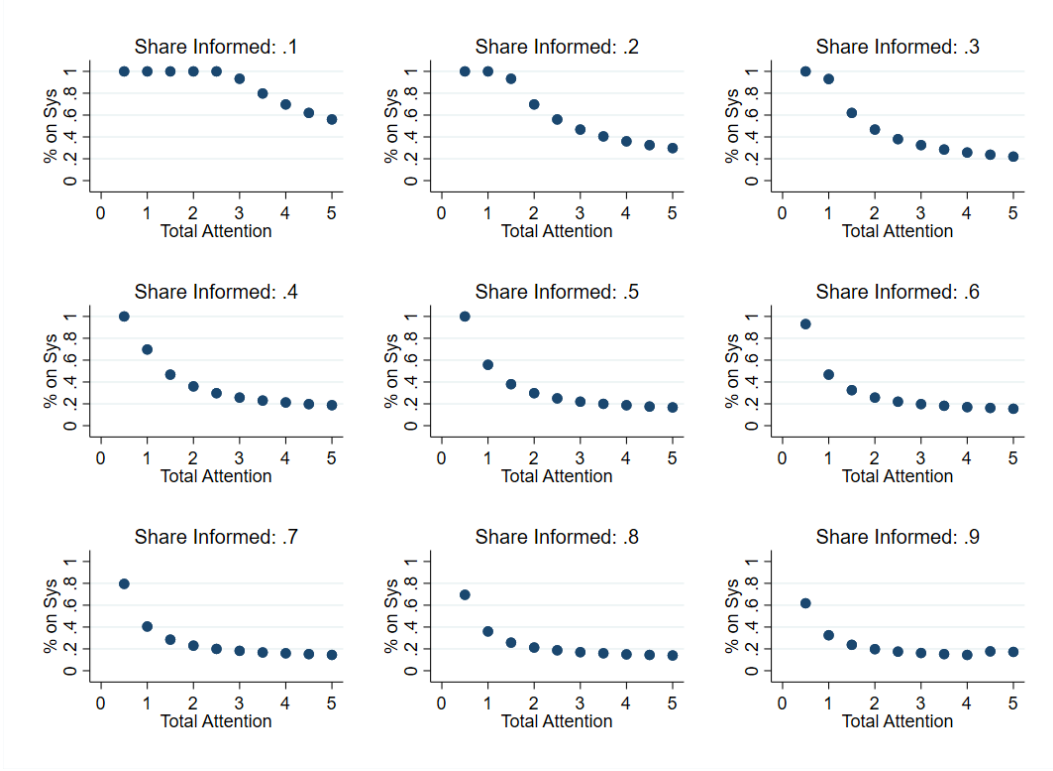
**Figure 17. Relationship between the size of the ETF and informed/uninformed investors' risk aversion.** The share of informed investors is set to 50%. The volatility of the systematic risk-factor  $\sigma_f = 0.25$ . If the ETF owned all the shares of the underlying stocks, it would have 20 shares outstanding.



**Figure 18. Relationship between risk aversion and attention to systematic risk-factor.** In the left panel,  $\sigma_n^2$  is set to 0.2, while in the right panel,  $\sigma_n^2$  is set to 0.5. In both panels,  $\rho^i$  is set high enough that the ETF is in zero average supply.

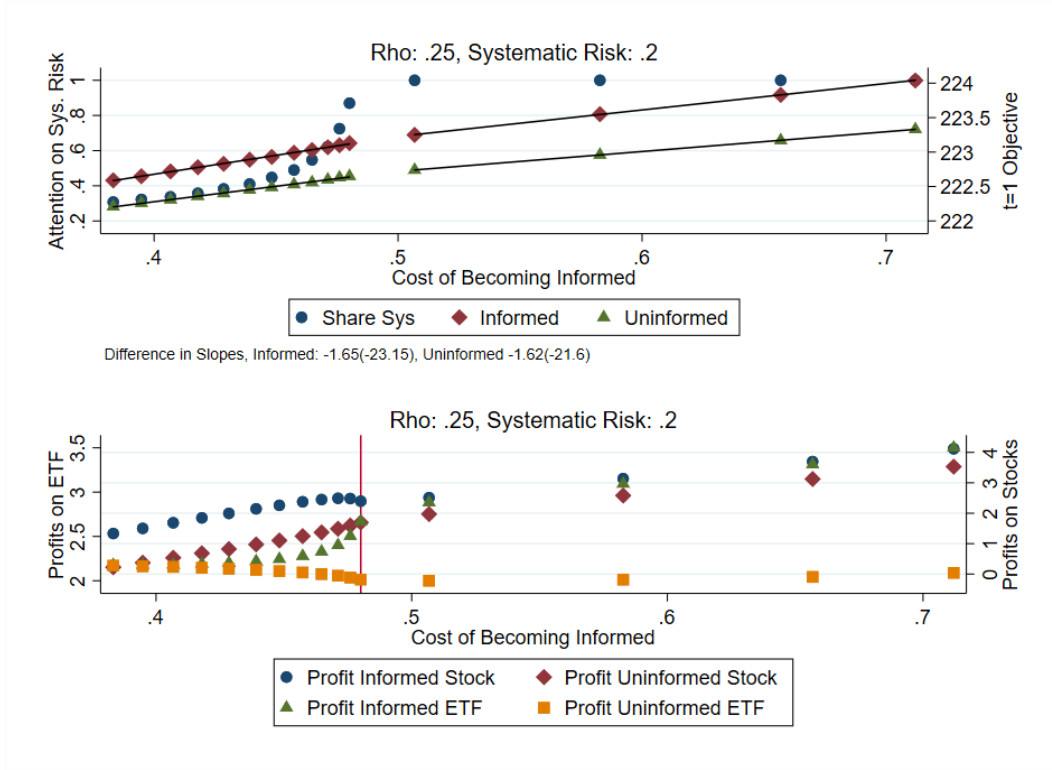


**Figure 19. Relationship between systematic risk and attention to systematic risk-factor.** In the left panel, risk aversion,  $\rho$  is set to 0.1, while in the right panel, risk aversion is set to 0.25. In both panels,  $\rho^i$  is set high enough that the ETF is in zero average supply.

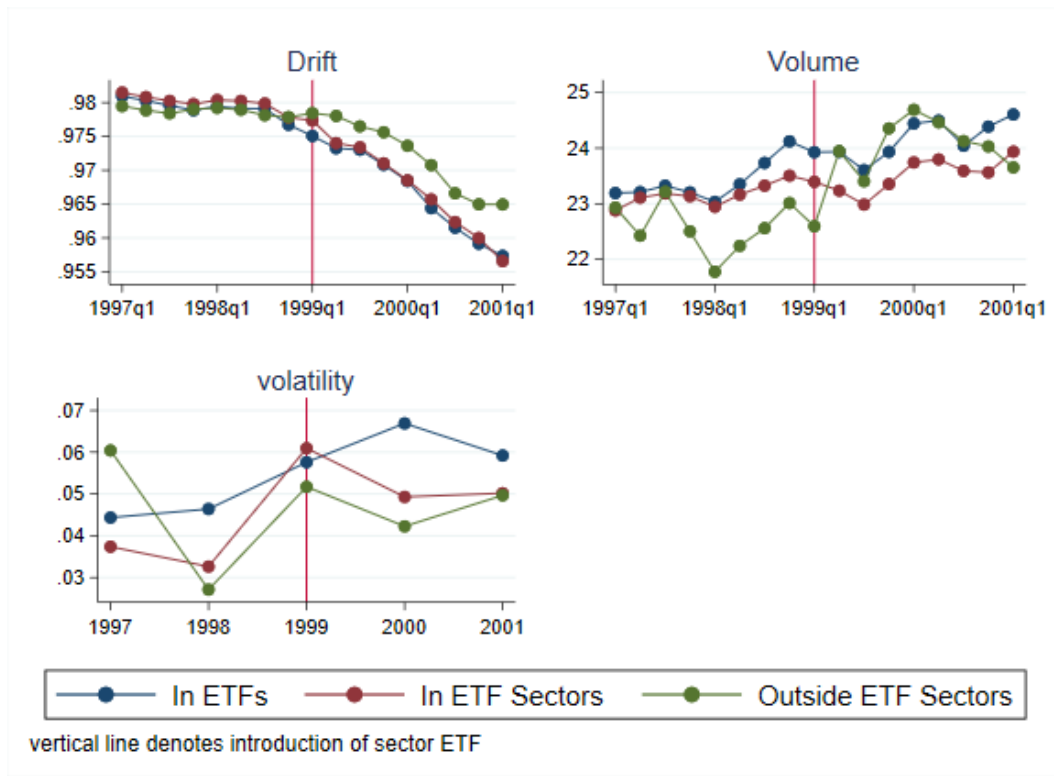


**Figure 20. Effect of Varying  $K$  and Share Informed.** x-axis is total attention  $K$ , y-axis is the share of total attention allocated to systematic risk. Parameters:  $\rho = 0.25$ ,  $\sigma_n = 0.25$ . ETF is present in zero average supply.

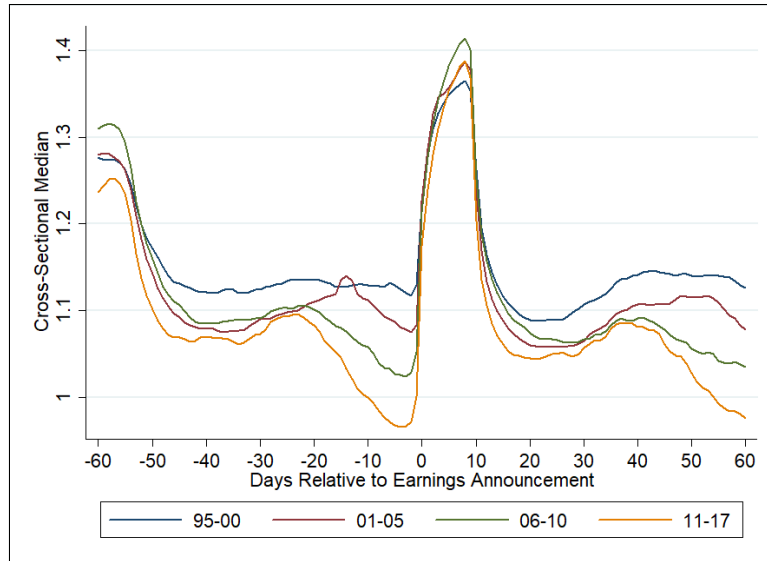




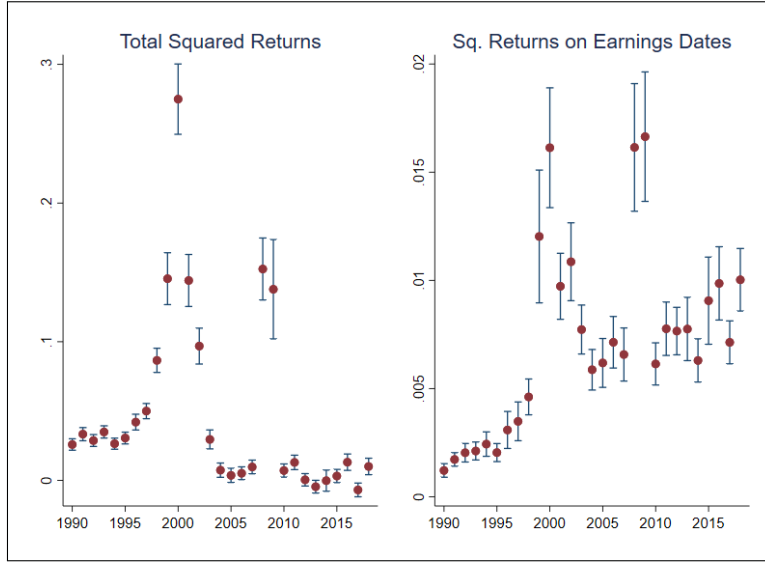
**Figure 21. Effect of larger  $\alpha$ .** Top panel: Effect of the cost of becoming informed on the share of attention on systematic risk. Bottom panel: Effect of the cost of becoming informed on trading profits.  $\alpha = 0.05$ . In both panels,  $\rho^i$  is set high enough that the ETF is in zero average supply.



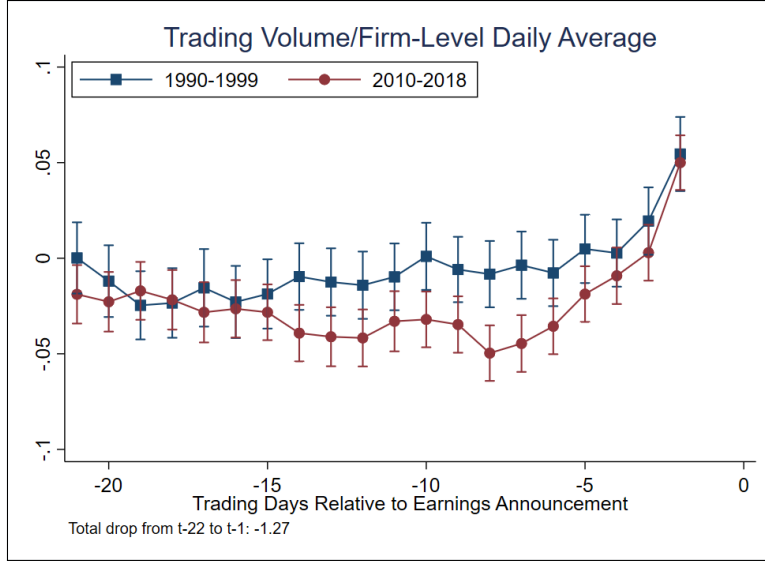
**Figure 22. Sector ETF Trends.** Trends in pre-earnings drift, pre-earnings trading volume and share of volatility on earnings days around the introduction of Sector ETFs by State Street in Q1 1999 (vertical red line).



**Figure 23. 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.



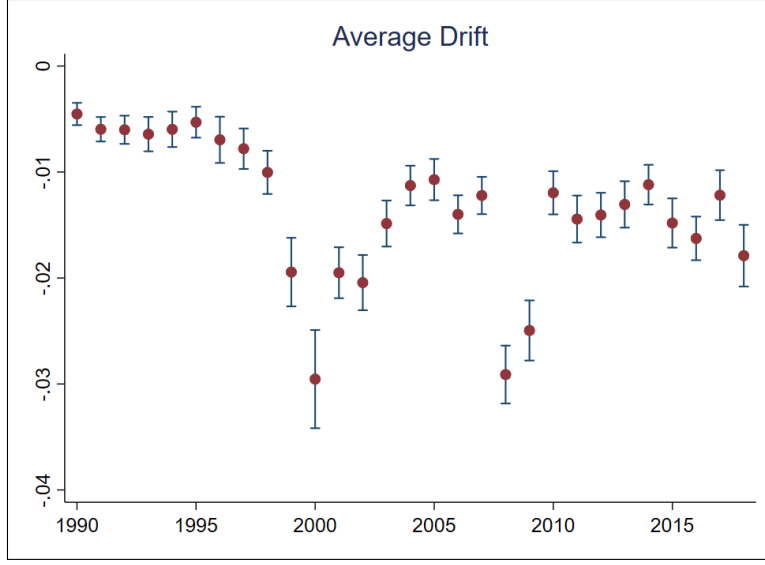
**Figure 24. Decomposition of QVS.** This figure plots coefficients from a regression of the numerator and denominator of QVS on a set of year dummy variables. For firm  $i$  in year  $t$  the quadratic variation share (QVS) is defined as:  $QVS_{i,t} = \frac{\sum_{\tau=1}^4 r_{i,\tau}^2}{\sum_{j=1}^{252} r_{i,j}^2}$ , where  $r$  denotes a market-adjusted daily return. The numerator (Sq. Returns on Earnings Dates) sums over the 4 quarterly earnings days in year  $t$ , while the denominator (Total Squared Returns) includes all days in calendar year  $t$ . Observations are weighted by 1-year lagged market capitalization. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.



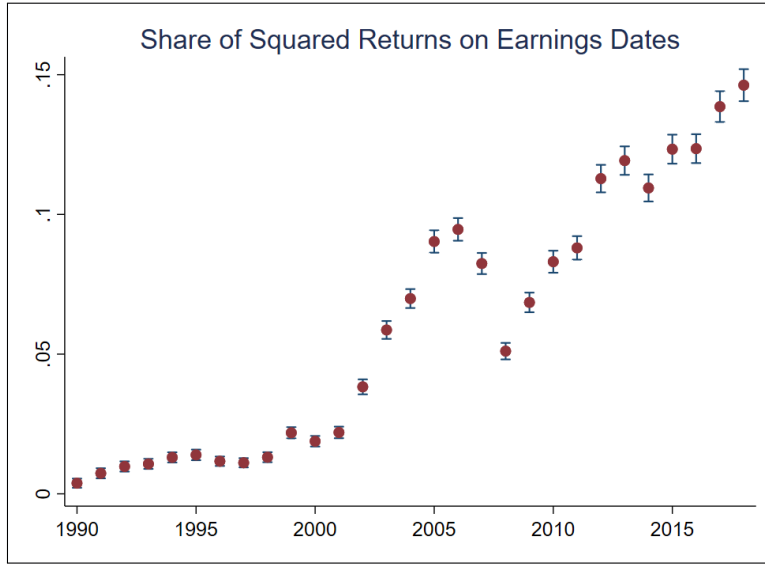
**Figure 25. Decline of Pre-Earnings Volume (Value Weighted).** Plot of  $\beta_k$  estimated from the regression:

$$AV_{i,t+\tau} = \alpha + \sum_{k=-21}^{22} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau}$$

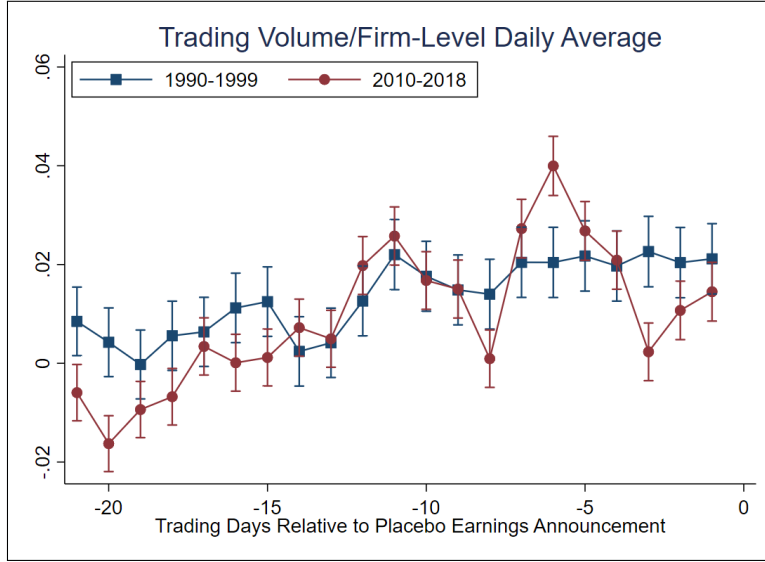
Bars represent a 95% confidence interval around the point estimate. Standard errors are clustered at the firm level. Observations are weighted by a firm's 1-year lagged market capitalization.



**Figure 26. Decline of Average Pre-Earnings Drift (Value Weighted).** This figure plots coefficients from a regression of pre-earnings drift on a set of year dummy variables where  $DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases}$ . A value near 1 implies most earnings information is incorporated in prices before the announcement date, while lower values denote less informative pre-earnings announcement prices. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level. Observations are weighted by a firm's 1-year lagged market capitalization.



**Figure 27. Increase in Earnings Day Volatility (Value Weighted).** This figure plots coefficients from a regression of QVS on a set of year dummy variables. 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  $t$ . Standard errors are clustered at the firm level. Observations are weighted by a firm's 1-year lagged market capitalization.

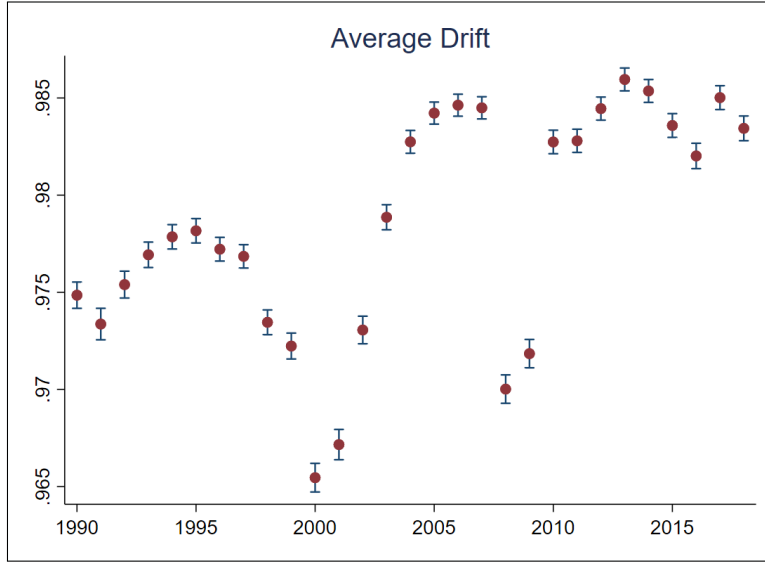


**Figure 28. Placebo Test: Pre-Earnings Volume.** Plot of  $\beta_k$  estimated from the regression:

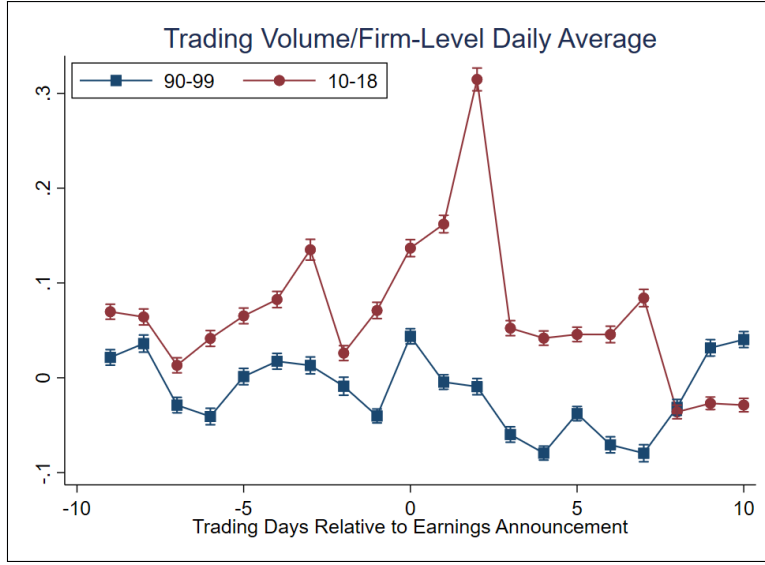
$$AV_{i,t+\tau} = \alpha + \sum_{k=-21}^{22} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau}$$

where  $t$  represents a placebo earnings date, 22 days before the actual earnings announcement date. Bars represent a 95% confidence interval around the point estimate. Standard errors are clustered at the firm level.





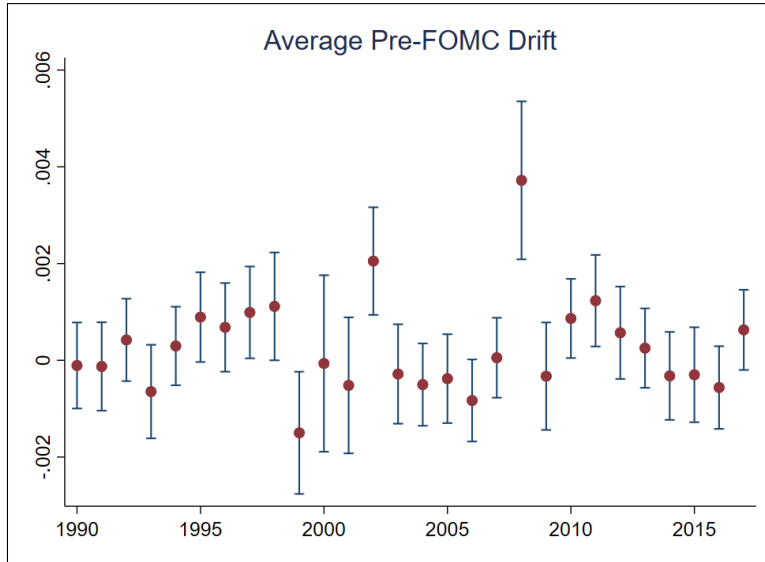
**Figure 29. Placebo Test: Pre-Earnings Drift.** This figure plots coefficients from a regression of pre-earnings drift on a set of year dummy variables where  $DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases}$  except  $t$  is a placebo earnings date 22 days before the actual earnings announcement date. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.



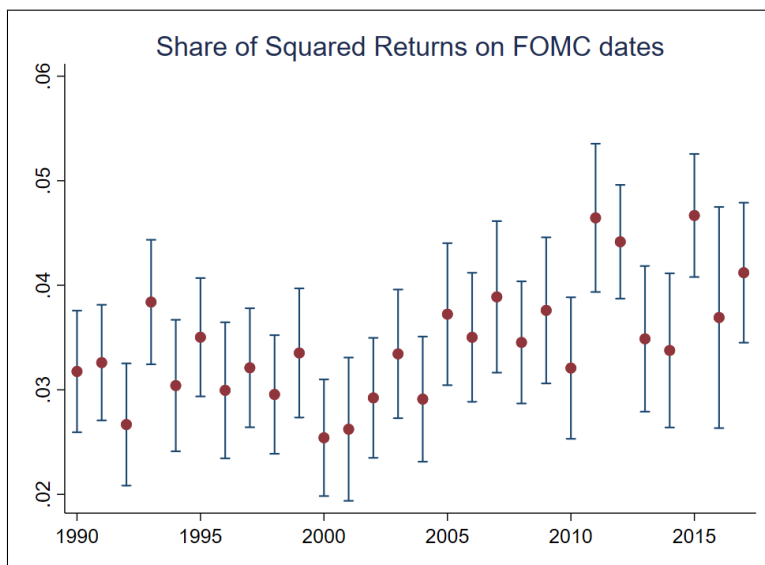
**Figure 30. FOMC Meeting Dates: Pre-Earnings Volume.** Plot of  $\beta_k$  estimated from the regression:

$$AV_{i,t+\tau} = \alpha + \sum_{k=-10}^{10} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau}$$

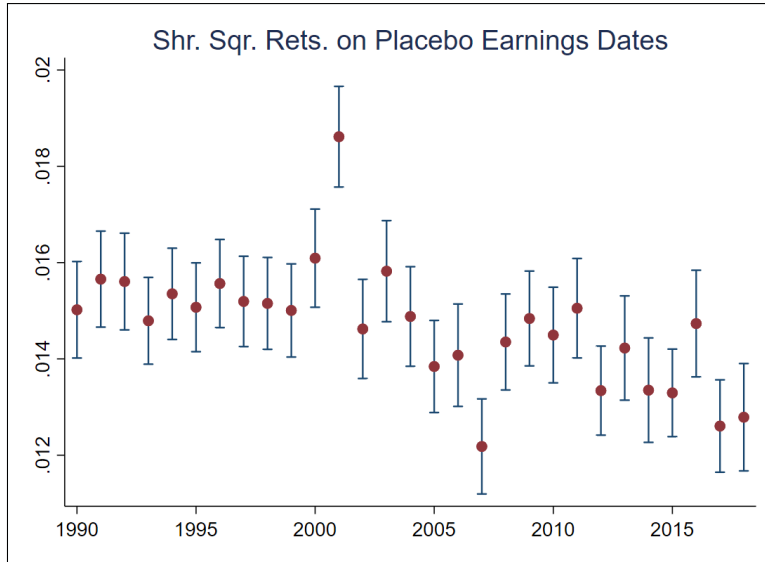
Bars represent a 95% confidence interval around the point estimate. Standard errors are clustered at the firm level.  $t$  represents a scheduled FOMC meeting date. The firm-level daily average is computed over the previous quarter. Observations are weighted by a firm's 1-year lagged market capitalization.



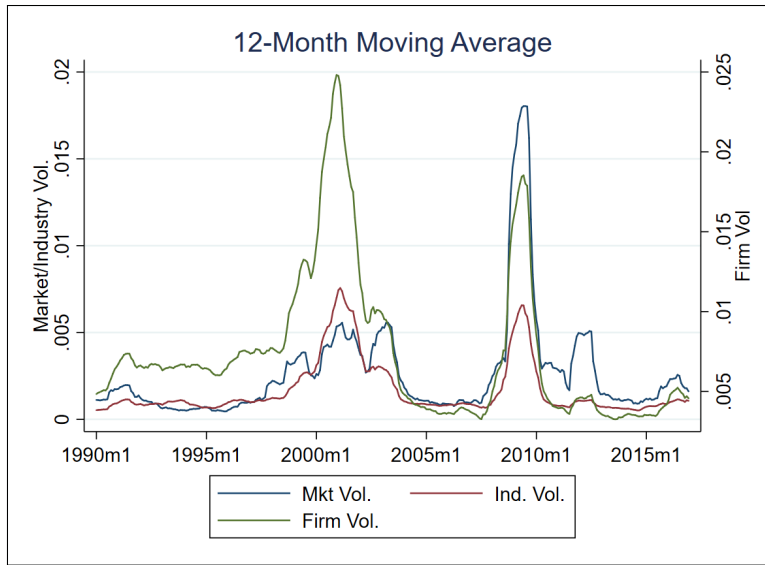
**Figure 31. FOMC Meeting Dates: Pre-Earnings Drift.** This figure plots coefficients from a regression of pre-earnings drift on a set of year dummy variables where  $DM_{it} = \begin{cases} \frac{1+r_{(t-10,t-1)}}{1+r_{(t-10,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-10,t)}}{1+r_{(t-10,t-1)}} & \text{if } r_t < 0 \end{cases}$ . Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.  $t$  denotes a scheduled FOMC meeting date. Observations are weighted by a firm's 1-year lagged market capitalization.



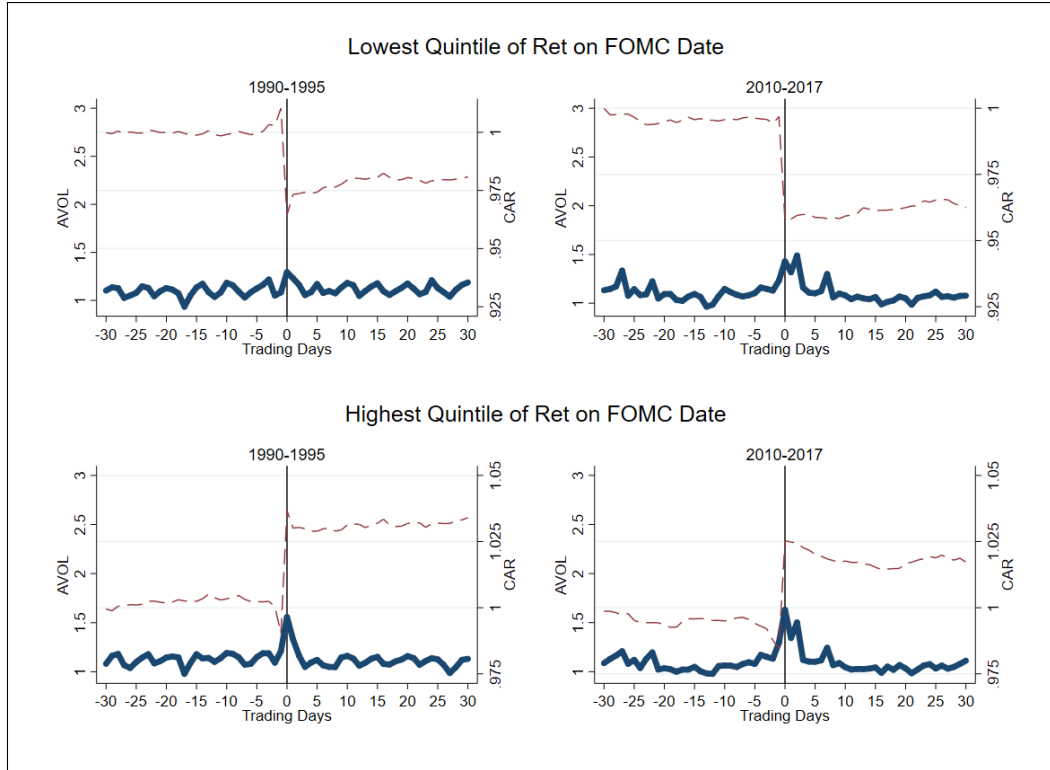
**Figure 32. FOMC Meeting Dates: Earnings Day Volatility.** This figure plots coefficients from a regression of QVS on a set of year dummy variables. For firm  $i$  in year  $t$  the quadratic variation share (QVS) is defined as:  $QVS_{i,t} = \sum_{\tau=1}^8 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , where  $r$  denotes a market-adjusted daily return. The numerator sums over the 8 scheduled FOMC meeting days in year  $t$ , while the denominator includes all days in calendar year  $t$ . Standard errors are clustered at the firm level. Observations are weighted by a firm's 1-year lagged market capitalization.



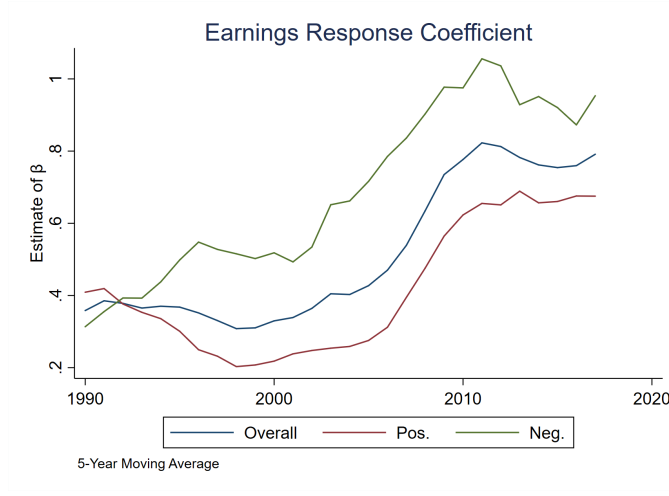
**Figure 33. Placebo Test: Earnings Day Volatility.** This figure plots the share of market-adjusted quadratic variation occurring on placebo earnings days. For firm  $i$  in year  $t$  the quadratic variation share (QVS) is defined as:  $QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , where  $r$  denotes a market-adjusted daily return. The numerator is the sum of squared returns on the 4 placebo earnings dates, each of which are 22 days before the actual earnings announcement dates, while the denominator is the sum of squared returns on all trading days in calendar year  $t$ .



**Figure 34. Replication and Expansion of Campbell et al. (2001).** This figure plots a 12-month moving average for the 3 components of total volatility. The market component in period  $t$  is  $\sum_{s \in t} (R_{ms} - \mu_m)^2$ , where  $R_{ms}$  is the return on the market at time  $s$  and  $\mu_m$  is the mean return on the market over the sample. Industry returns are assumed to follow  $R_{it} = R_{mt} + \epsilon_{it}$ , so the industry component is the sum of squared  $\epsilon_{it}$ 's. The firm component is any residual volatility not explained by the industry and market components.



**Figure 35. Returns and Trading Volume around FOMC meeting dates.** Each plot represents the cross-sectional average for each quintile of Return on FOMC dates over the specified years. Abnormal returns are defined as returns relative to the returns on the CRSP value-weighted index, as in Campbell et al. (2001). Cumulative abnormal returns (CAR) are calculated by compounding net abnormal returns. Cumulative abnormal volume is volume relative to the firm-level average over the past year, see Equation 23.



**Figure 36. Trends in Earnings Response.** This table contains the results of the following regression:

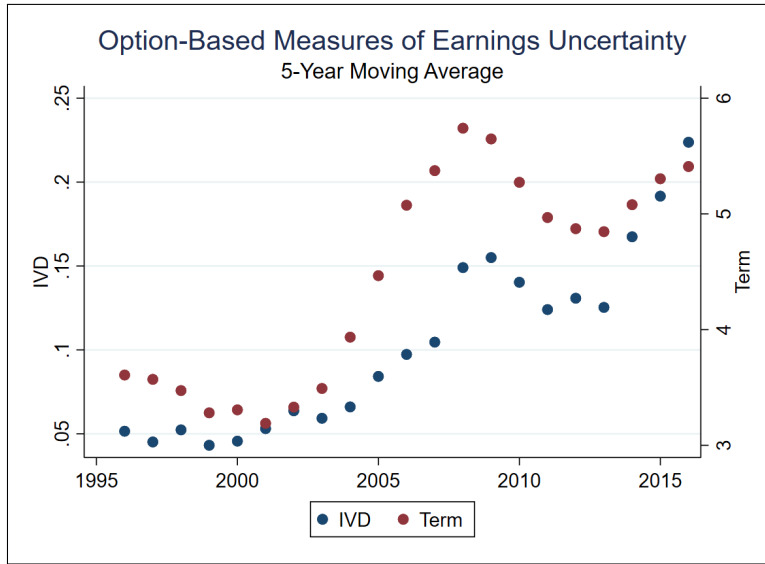
$$r_{i,t} = \alpha + \beta \times SUE_{i,t} + controls + \epsilon_{i,t}$$

and

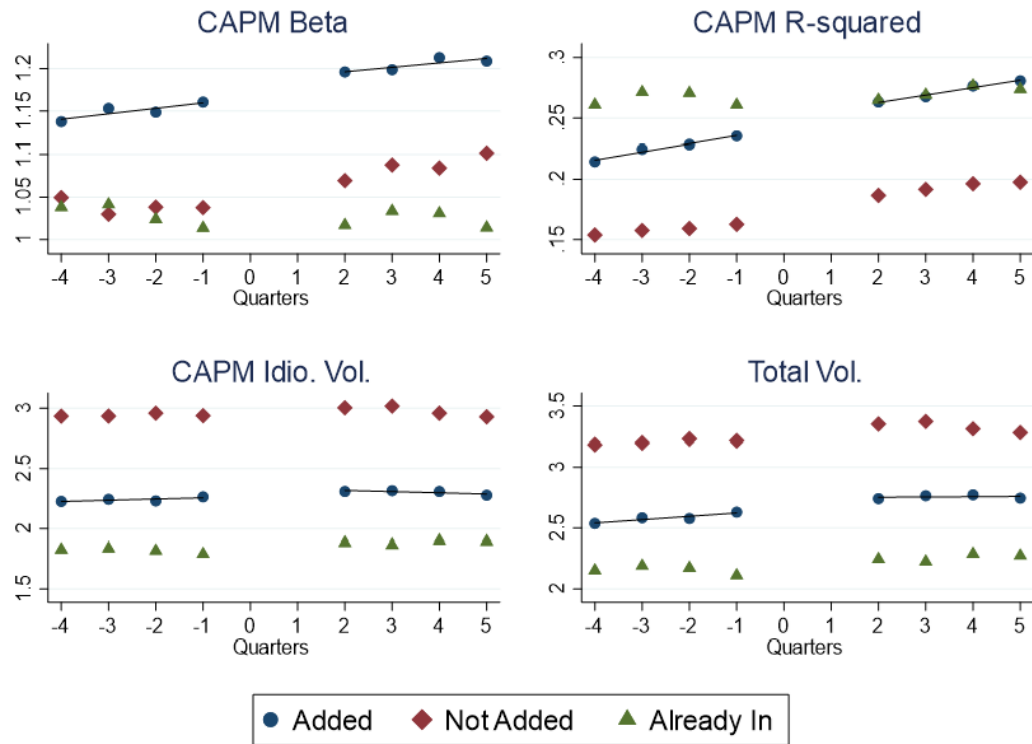
$$r_{i,t} = \alpha + \beta_1 \times SUE_{i,t} \times \mathbf{1}_{SUE_{i,t} > 0} + \beta_2 \times |SUE_{i,t}| \times \mathbf{1}_{SUE_{i,t} < 0} + controls + \epsilon_{i,t}$$

Here,  $r_{i,t}$  denotes the market-adjusted return on the effective quarterly earnings date. *Overall* is coefficient from baseline earnings-response regression. *Pos.* and *Neg.* are coefficients from the earnings-response regression which allows for asymmetric effects of positive and negative earnings surprises.

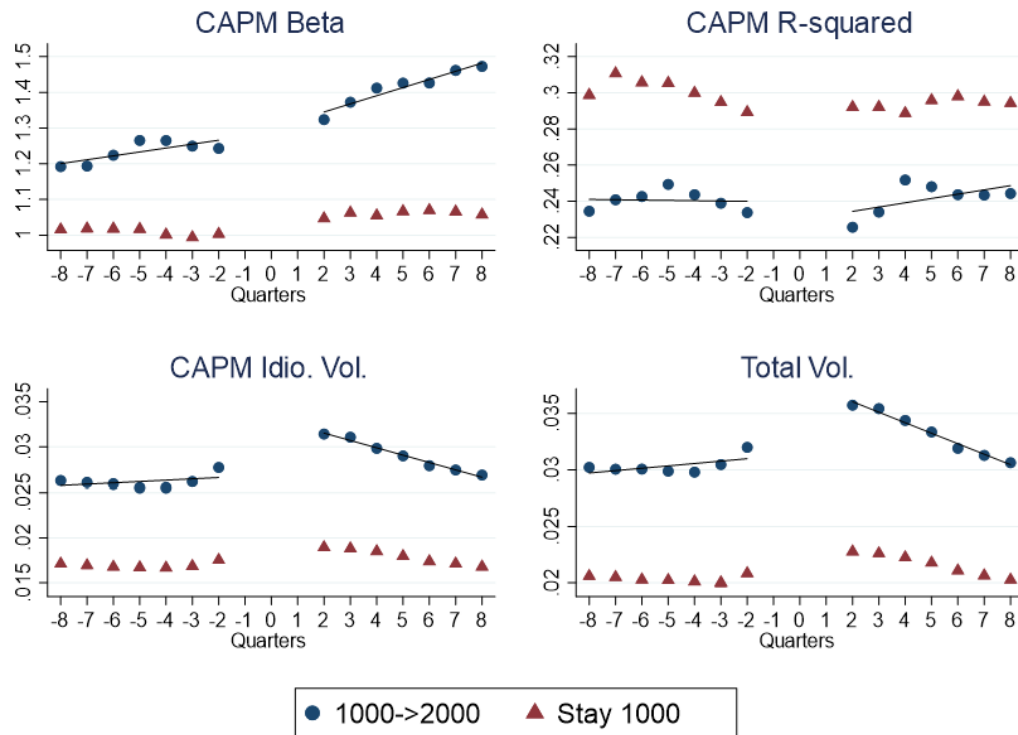




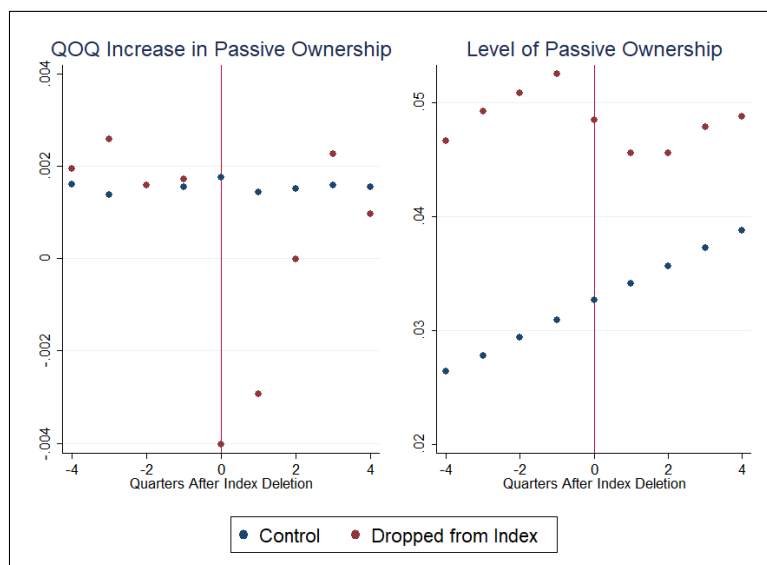
**Figure 37. Option-Based Measures of Earnings Uncertainty.** Each point represents the cross-sectional average of each option-based measure of earnings uncertainty. For earnings date  $\tau$ ,  $IVD = \frac{\overline{IV}_b}{0.5(\overline{IV}_a + \overline{IV}_c)}$  and  $Term = \frac{(\sigma_{term}^Q)^2}{IV_b}$ . Sample includes 306 S&P 500 firms with options that meet the filters described in Kelly et al. (2016), and have at least 16 years with 4 non-missing earnings announcements.



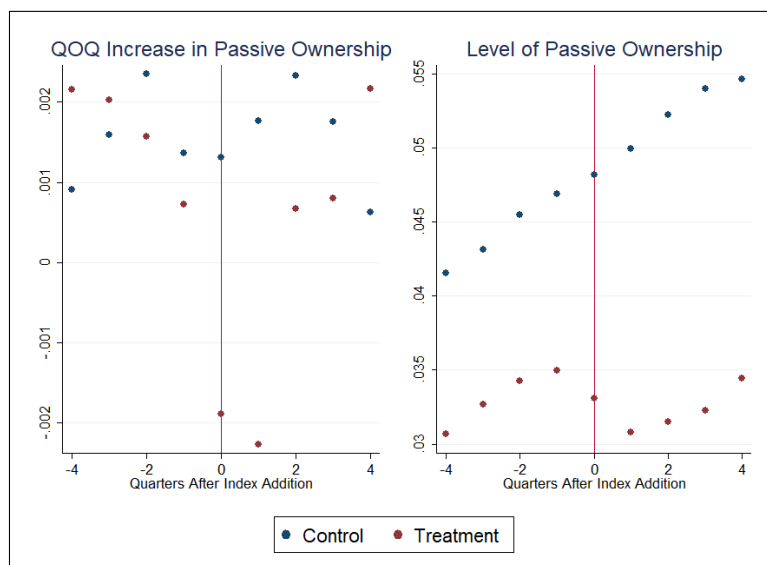
**Figure 38. Effect of S&P 500 Index Addition on CAPM  $\beta$ , CAPM  $R^2$ , Idiosyncratic Volatility and Total Volatility.** Quarters are quarters relative to index addition.



**Figure 39.** Russell 1000/2000 rebalancing on CAPM  $\beta$ , CAPM  $R^2$ , Idiosyncratic Volatility and Total Volatility. Quarters are quarters relative to index rebalancing.



**Figure 40. 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 41. 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.

## List of Tables

$t = 0$	•	<ul style="list-style-type: none"> <li>◦ Binary decision: pay fixed-cost <math>c</math> and become informed or stay uninformed</li> <li>◦ Informed allocate <math>K</math> units of attention to the underlying risks</li> </ul>
$t = 1$	•	<ul style="list-style-type: none"> <li>◦ Informed investors receive signals about asset payoffs</li> <li>◦ Informed/ uninformed submit demands</li> </ul>
$t = 2$	•	Payoffs realized, investors consume

**Table 1 Model Timeline.**

$\rho$	$\sigma_f^2$	Shr. Inf.	$G_{i,i}$	No ETF		ETF	
				$G_{i,j}$	$7 \times G_{i,j}$	$G_{i,i}$	$G_{i,j}$
0.1	0.2	0.5	0.968	-0.117	-0.817	1.260	-1.260
0.1	0.5	0.5	0.766	-0.069	-0.484	1.010	-1.010
0.25	0.2	0.5	0.290	-0.024	-0.171	0.274	-0.274
0.25	0.5	0.5	0.255	-0.019	-0.130	0.124	-0.124
0.35	0.2	0.5	0.189	-0.014	-0.100	0.046	-0.046
0.35	0.5	0.5	0.176	-0.012	-0.086	0.003	-0.003

**Table 2 Hedging Demand.** The share of informed investors are fixed and 60%. The “No ETF” columns are the entries of  $G_1$  when the ETF is not present, while the “ETF” column is the entries of  $G_1$  after introducing the ETF in zero average supply. There are 8 assets in the economy, so  $7 \times G_{i,j}$  is the total hedging of systematic risk when betting on a stock-specific signal when the ETF is not present.



Panel A: Intensive Margin Effects									
	No ETF	Drift High $\rho^i$	Low $\rho^i$	No ETF	Volatility High $\rho^i$	Low $\rho^i$	No ETF	Volume High $\rho^i$	Low $\rho^i$
Attention to Idio. Risk	0.263*** (0.019)	0.0994*** (0.009)	0.0950*** (0.008)	-3.577*** (0.346)	-1.736*** (0.154)	-1.776*** (0.149)	23.78*** (2.781)	12.00*** (1.282)	12.27*** (1.235)
Panel B: Extensive Margin Effects									
	No ETF	Drift High $\rho^i$	Low $\rho^i$	No ETF	Volatility High $\rho^i$	Low $\rho^i$	No ETF	Volume High $\rho^i$	Low $\rho^i$
Share Informed	-0.145*** (0.023)	-0.0114 (0.008)	-0.00988 (0.008)	0.212 (0.417)	-0.415*** (0.136)	-0.496*** (0.146)	-11.31*** (3.596)	-0.0532 (1.185)	0.465 (1.220)

**Table 3 Effect of Learning on Price Informativeness.** Panel A is a regression of each of the price informativeness measures on the share of attention devoted to stock-specific risks. This regression includes fixed effects for the share of informed agents. Panel B is a regression of each of the price informativeness measures on the share of informed agents. This regression includes fixed effects for investor risk aversion  $\rho$  and the volatility of the systematic risk factor  $\sigma_n$ . It also includes the share of attention devoted to idiosyncratic risk as a right-hand-side variable. The “No ETF” columns are from economies where the ETF is not present, while the ETF is present in the “High  $\rho^i = 100$ ” and “Low  $\rho^i = 0$ ” columns.

	(1)	(2)	(3)
Inc. Passive	-12.81*** (1.986)	-16.09*** (2.487)	-23.96*** (5.416)
Observations	239,859	239,859	239,859
R-squared	0.022	0.04	0.112
Controls	No	Yes	Yes
Firm FE	No	Yes	Yes
Weight	Eq.	Eq.	Val.

**Table 4 Passive Ownership and Pre-Earnings Volume.** Estimates of  $\beta$  from:

$$\Delta CAV_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$  is cumulative abnormal pre-earnings trading volume.  $\Delta$  is a year-over-year change, matching on fiscal quarter. Change in passive ownership is expressed as a decimal, so  $0.01 = 1\%$  increase. Controls,  $X_{i,t-1}$ , include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, total institutional ownership and growth in market capitalization from  $t - 1$  to  $t$ . Standard errors are computed using panel Newey-West with 8 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter.

	(1)	(2)	(3)
Inc. Passive	-0.0298** (0.012)	-0.0322** (0.014)	-0.0965*** (0.029)
Observations	239,689	239,689	239,689
R-squared	0.02	0.045	0.063
Controls	No	Yes	Yes
Firm FE	No	Yes	Yes
Weight	Eq.	Eq.	Val.

**Table 5 Passive Ownership and Pre-Earnings Drift.** Table with estimates of  $\beta$  from:

$$\Delta DM_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

Where  $DM_{i,t}$  is a measure of the pre-earnings drift. Passive ownership is expressed as a decimal, so  $0.01 = 1\%$  of shares outstanding held by passive funds. Controls,  $X_{i,t-1}$ , include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, total institutional ownership and growth in market capitalization from  $t - 1$  to  $t$ . Standard errors are computed using panel Newey-West with 8 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter.

	(1)	(2)	(3)
Inc. Passive	0.200*** (0.031)	0.106*** (0.036)	0.381** (0.180)
Observations	127,951	126,319	126,319
R-squared	0.011	0.03	0.035
Controls	No	Yes	Yes
Firm FE	No	Yes	Yes
Weight	Eq.	Eq.	Val.

**Table 6 Passive Ownership and Earnings Day Share of Volatility.** Table with estimates of  $\beta$  from:

$$\Delta QVS_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , 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  $t$ . Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . Fixed effects are year and firm. Standard errors are computed using panel Newey-West with 2 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-year lagged market capitalization relative to other firms that year.

		Volume	Drift	Volatility
Model	No ETF	0.9898	0.96388	0.70028
	ETF 0%	0.94386	0.95526	0.7099
	ETF 15%	0.92784	0.95518	0.71903
	Change	-0.062	-0.0087	0.01874
Data	15% $\Delta$ Passive	-3.594	-0.0145	0.05715

**Table 7 Calibrating the Model to Match the Reduced-Form Results.** *The cost of becoming informed is set so 60% learn in equilibrium when the ETF is not present,  $\rho = 0.15$ , and  $\sigma_f^2=0.25$ . To match the average passive ownership of 15% in 2018  $\rho^i$  is set to 1. Data row is effect of a 15% increase in passive ownership based on value-weighted reduced-form estimates.*

	# Analysts	Distance	Time Between Updates	Downloads
Inc. Passive	-8.935*** (0.824)	1.557*** (0.244)	14.93* (8.692)	-3.756*** (1.185)
Observations	99,004	96,365	79,131	96,380
R-squared	0.1	0.062	0.065	0.233
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Weight	Eq.	Eq.	Eq.	Eq.

**Table 8 Investor Attention and Passive Ownership.** This table contains the estimates of  $\beta$  from:

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

Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . Fixed effects include year/quarter and firm.

	Volume	Drift	Volatility
Inc. Passive	-51.08** (22.550)	-0.322** (0.140)	1.924** (0.768)
R-squared	0.098	0.074	0.115
Reduced Form	-23.96***	-0.0965***	0.381**
Treated firms	419	419	419
Control (out)	906	906	906
Control (in)	508	508	508

**Table 9 Effects of S&P 500 Index Addition.** Two-stage least squares estimates from first stage:

$$\Delta Passive_{i,t} = \alpha + \beta \times Treated_{i,t} + \gamma_t + \epsilon_{i,t}$$

Second stage:

$$\Delta Outcome_{i,t} = \alpha + \beta \times \widehat{\Delta Passive}_{i,t} + \gamma_t + \epsilon_{i,t}$$

Where  $Treated_{i,t}$  is a dummy variable equal to one if a firm was added to the S&P 500 index, and  $\gamma_t$  is a month of index addition fixed effect.

	Volume	Drift	Volatility
Inc. Passive	-44.71** (20.740)	-0.285** (0.125)	0.0109 (0.411)
R-squared	0.099	0.126	0.073
Reduced Form	-23.96***	-0.0965***	0.381**

**Table 10 Effects of Russell 1000/2000 Index Reconstitution.** Two-stage least squares estimates from first stage:

$$\Delta Passive_{i,t} = \alpha + \beta \times Treated_{i,t} + \gamma_t + \epsilon_{i,t}$$

Second stage:

$$\Delta Outcome_{i,t} = \alpha + \beta \times \widehat{\Delta Passive}_{i,t} + \gamma_t + \epsilon_{i,t}$$

Where  $Treated_{i,t}$  is a dummy variable equal to one if a firm was added to the S&P 500 index, and  $\gamma_t$  is a month of index addition fixed effect. There are 216 treated firms and 158 control firms.



$t = 0$	<ul style="list-style-type: none"> <li>◦ Binary decision: pay fixed-cost <math>c</math> and become informed or stay uninformed</li> <li>◦ Informed allocate one unit of attention to the underlying risks</li> <li>◦ <b>Intermediary submits market order to buy <math>v</math> shares of the each stock.</b></li> </ul>
$t = 1$	<ul style="list-style-type: none"> <li>◦ <b>Intermediary's market order clears, leaving <math>\bar{x} - v</math> shares of each stock, and <math>v \times n</math> shares of the ETF available for purchase</b></li> <li>◦ Informed investors receive private signals</li> <li>◦ Informed/ uninformed submit demands</li> </ul>
$t = 2$	<ul style="list-style-type: none"> <li>◦ Payoffs realized, investors consume</li> </ul>

**Table 11New Model Timeline.** Differences from original timeline in bold.

Mean asset payoff	$a_i$	15
Volatility of idiosyncratic shocks	$\sigma_i^2$	0.55
Volatility of noise shocks	$\sigma_x^2$	0.5
Risk-free rate	$r$	1
Initial wealth	$w_0$	220
Baseline Learning	$\alpha$	0.001
# idiosyncratic assets	$n$	8
Coef. of risk aversion (low)	$\rho$	0.1
Coef. of risk aversion (high)	$\rho$	0.35
Vol. of systematic shocks (low)	$\sigma_n^2$	0.2
Vol. of systematic shocks (high)	$\sigma_n^2$	0.5
Total supply of idiosyncratic assets	$\bar{x}$	20

**Table 12 Baseline Parameters.**

Panel A: Matching Cost of Becoming Informed							
$\rho$	$\sigma_n^2$	Precision					
		Share Informed		Informed		Uninformed	
		no ETF	ETF	no ETF	ETF	no ETF	ETF
0.1	0.2	0.05	0.2	1.82	2.24	1.66	2.06
0.1	0.5	0.35	0.2	2.04	2.06	1.93	1.94
0.25	0.2	0.5	0.2	1.85	1.87	1.74	1.82
0.25	0.5	0.5	0.2	1.78	1.87	1.69	1.82
Panel B: Share Informed at 10%							
$\rho$	$\sigma_n^2$	Precision					
		Share Informed		Informed		Uninformed	
		no ETF	ETF	no ETF	ETF	no ETF	ETF
0.1	0.2	0.1	0.1	1.85	2.05	1.70	1.88
0.1	0.5	0.1	0.1	1.75	1.90	1.64	1.83
0.25	0.2	0.1	0.1	1.76	1.87	1.65	1.82
0.25	0.5	0.1	0.1	1.71	1.87	1.62	1.82
Panel C: Share Informed at 30%							
$\rho$	$\sigma_n^2$	Precision					
		Share Informed		Informed		Uninformed	
		no ETF	ETF	no ETF	ETF	no ETF	ETF
0.1	0.2	0.3	0.3	2.20	2.54	2.05	2.37
0.1	0.5	0.3	0.3	1.96	2.30	1.85	2.16
0.25	0.2	0.3	0.3	1.79	1.92	1.68	1.84
0.25	0.5	0.3	0.3	1.73	1.88	1.64	1.83

**Table 13 Posterior Precision.** Diagonal entries of  $\hat{\Sigma}$  for one of the stocks i.e. assets 1 to  $n$ . In panel A, the cost of being informed is chosen such that 20% of investors become informed when the ETF is present. In Panels B and C, the share of informed investors are fixed and 10% and 30% respectively. The “no ETF” column has the (1,1) entry of  $\hat{\Sigma}$  when the ETF is not present, while the “ETF” column has the (1,1) entry of  $\hat{\Sigma}$  after introducing the ETF. In the “ETF” column, the ETF is in zero average supply.

Panel A: Fix Share Informed

$\rho$	$\sigma_f^2$	Shr. Inf.	Risk Premium		Change(PP)
			No ETF	ETF	
0.1	0.2	0.1	3.73%	3.71%	-0.02%
0.1	0.2	0.3	3.71%	3.59%	-0.12%
0.1	0.5	0.1	8.18%	8.19%	0.01%
0.1	0.5	0.3	8.09%	8.05%	-0.04%
0.35	0.2	0.1	14.33%	14.32%	-0.01%
0.35	0.2	0.3	14.28%	14.23%	-0.05%
0.35	0.5	0.1	35.98%	36.09%	0.11%
0.35	0.5	0.3	35.65%	35.94%	0.30%

Panel B: Fix Cost of Becoming Informed

$\rho$	$\sigma_f^2$	Risk Premium		Change(PP)
		No ETF	ETF	
0.1	0.2	3.68%	3.38%	-0.30%
0.1	0.5	7.98%	8.19%	0.21%
0.35	0.2	14.23%	14.23%	0.00%
0.35	0.5	35.32%	35.94%	0.63%

**Table 14 Effect of introducing the ETF on Expected Returns.** In Panel A, the share informed is the same whether the ETF is present or not. In Panel B, the share informed when the ETF is not present is set to 50%. After introducing the ETF, the share informed are 0.55, 0.2, 0.3 and 0.3 in rows 1-4. The risk premium is defined as the average stock return between period 0 and period 2. When the ETF is present, it is in zero average supply.

Panel A: Matching Cost of Becoming Informed					
$\rho$	$\sigma_n^2$	Share Informed		Diff. in EU	
		no ETF	ETF	no ETF	ETF
0.1	0.2	0.05	0.2	0.154%	0.163%
0.1	0.5	0.35	0.2	0.181%	0.177%
0.25	0.2	0.5	0.2	0.229%	0.229%
0.25	0.5	0.5	0.2	0.572%	0.571%

Panel B: Share Informed at 10%					
$\rho$	$\sigma_n^2$	Share Informed		Diff. in EU	
		no ETF	ETF	no ETF	ETF
0.1	0.2	0.1	0.1	0.154%	0.177%
0.1	0.5	0.1	0.1	0.226%	0.186%
0.25	0.2	0.1	0.1	0.251%	0.296%
0.25	0.5	0.1	0.1	0.727%	1.103%

Panel C: Share Informed at 30%					
$\rho$	$\sigma_n^2$	Share Informed		Diff. in EU	
		no ETF	ETF	no ETF	ETF
0.1	0.2	0.3	0.3	0.132%	0.141%
0.1	0.5	0.3	0.3	0.190%	0.154%
0.25	0.2	0.3	0.3	0.237%	0.211%
0.25	0.5	0.3	0.3	0.650%	0.300%

**Table 15 Effect of Introducing the ETF on Expected Utility of Informed and Uninformed.** This table quantifies the effect of introducing the ETF on the expected utility of informed and uninformed investors. The columns of interest are under the header “Diff. in EU”. The “no ETF” column is the % difference in expected utility between informed and uninformed investors when the ETF is not present. The ETF column repeats this exercise after introducing the ETF in zero average supply.

$\rho$	$\sigma_n^2$	Uninformed						
		Share Informed		No ETF Present		ETF Present		
		no ETF	ETF	Own	Stock Hedge	Own	Stock Hedge	ETF Hedge
0.1	0.2	0.05	0.2	6.333	-0.278	2.273	0.000	-2.273
0.1	0.5	0.35	0.2	1.764	-0.170	3.082	0.000	-3.082
0.25	0.2	0.5	0.2	2.380	-0.181	5.510	0.000	-5.510
0.25	0.5	0.5	0.2	2.550	-0.291	5.510	0.000	-5.510

$\rho$	$\sigma_n^2$	Informed						
		Share Informed		No ETF Present		ETF Present		
		no ETF	ETF	Own	Stock Hedge	Own	Stock Hedge	ETF Hedge
0.1	0.2	0.1	0.2	7.872	-0.489	4.023	0.000	-4.023
0.1	0.5	0.35	0.2	2.865	-0.270	4.307	0.000	-4.307
0.25	0.2	0.5	0.2	2.803	-0.218	5.710	0.000	-5.710
0.25	0.5	0.5	0.2	2.913	-0.317	5.710	0.000	-5.710

**Table 16 Sensitivity of Demand to Prices (fixed  $c$ ).** Entries of  $G_{2,inf}$  and  $G_{2,un}$  for one of the stocks i.e. assets 1 to  $n - 1$ . The cost of being informed is chosen such that 20% of investors become informed when the ETF is present. The “Own” columns are diagonal entries e.g. (1,1). The “Stock Hedge” column is one of the edge entries excluding the  $n^{th}$  e.g. (1,2) or (2,1). The “ETF Hedge” column is the  $n^{th}$  edge entry. ETF is present in zero average supply.

Panel A: Share Informed Fixed at 10%								
Uninformed								
$\rho$	$\sigma_n^2$	Share Informed		No ETF Present		ETF Present		
		no ETF	ETF	Own	Stock Hedge	Own	Stock Hedge	ETF Hedge
0.1	0.2	0.1	0.1	4.096	-0.036	4.040	0.000	-4.040
0.1	0.5	0.1	0.1	4.899	-0.528	7.656	0.000	-7.656
0.25	0.2	0.1	0.1	4.884	-0.464	6.270	0.000	-6.270
0.25	0.5	0.1	0.1	4.976	-0.601	6.270	0.000	-6.270
Informed								
$\rho$	$\sigma_n^2$	Share Informed		No ETF Present		ETF Present		
		no ETF	ETF	Own	Stock Hedge	Own	Stock Hedge	ETF Hedge
0.1	0.2	0.1	0.1	5.597	-0.236	5.790	0.000	-5.790
0.1	0.5	0.1	0.1	5.979	-0.623	8.343	0.000	-8.343
0.25	0.2	0.1	0.1	5.299	-0.499	6.470	0.000	-6.470
0.25	0.5	0.1	0.1	5.331	-0.626	6.470	0.000	-6.470
Panel B: Share Informed Fixed at 30%								
Uninformed								
$\rho$	$\sigma_n^2$	Share Informed		No ETF Present		ETF Present		
		no ETF	ETF	Own	Stock Hedge	Own	Stock Hedge	ETF Hedge
0.1	0.2	0.3	0.3	1.774	0.059	1.581	0.000	-1.581
0.1	0.5	0.3	0.3	2.020	-0.197	1.950	0.000	-1.950
0.25	0.2	0.3	0.3	3.190	-0.266	4.018	0.000	-4.018
0.25	0.5	0.3	0.3	3.364	-0.393	4.914	0.000	-4.914
Informed								
$\rho$	$\sigma_n^2$	Share Informed		No ETF Present		ETF Present		
		no ETF	ETF	Own	Stock Hedge	Own	Stock Hedge	ETF Hedge
0.1	0.2	0.3	0.3	3.198	-0.117	3.331	0.000	-3.331
0.1	0.5	0.3	0.3	3.121	-0.296	3.337	0.000	-3.337
0.25	0.2	0.3	0.3	3.614	-0.303	4.356	0.000	-4.356
0.25	0.5	0.3	0.3	3.728	-0.419	5.114	0.000	-5.114

**Table 17 Sensitivity of Demand to Prices (fixed share informed).** Entries of  $G_{2,inf}$  and  $G_{2,un}$  for one of the stocks i.e. assets 1 to  $n - 1$ . In Panels A and B, the share of informed investors are fixed and 10% and 30% respectively. The “Own” columns are diagonal entries e.g. (1,1). The “Stock Hedge” column is one of the edge entries excluding the  $n^{th}$  e.g. (1,2) or (2,1). The “ETF Hedge” column is the  $n^{th}$  edge entry. ETF is present in zero average supply.

Name	Ticker	Founded	Expense Ratio	NAV	AUM
The Consumer Discretionary Select Sector Fund	XLY	3/31/1999	0.13%	\$132.26	\$14,012.89 M
The Consumer Staples Select Sector Fund	XLP	3/31/1999	0.13%	\$60.54	\$14,241.29 M
The Energy Select Sector Fund	XLE	3/31/1999	0.13%	\$45.09	\$12,164.02 M
The Financial Select Sector Fund	XLF	3/31/1999	0.13%	\$26.16	\$21,849.83 M
The Health Care Select Sector Fund	XLV	3/31/1999	0.13%	\$102.82	\$26,580.85 M
The Industrial Select Sector Fund	XLI	3/31/1999	0.13%	\$74.39	\$10,293.96 M
The Materials Select Sector Fund	XLB	3/31/1999	0.13%	\$59.18	\$5,209.07 M
The Technology Select Sector Fund	XLK	3/31/1999	0.13%	\$102.42	\$30,947.40 M
The Utilities Select Sector Fund	XLU	3/31/1999	0.13%	\$61.55	\$11,819.93 M
Bank ETF	KBE	12/30/2005	0.35%	\$36.90	\$1,666.27 M
Capital Markets ETF	KCE	12/30/2005	0.35%	\$59.82	\$25.42 M
Insurance ETF	KIE	12/30/2005	0.35%	\$30.36	\$629.94 M
Biotech ETF	XBI	3/31/2006	0.35%	\$104.56	\$4,577.23 M
Homebuilders ETF	XHB	3/31/2006	0.35%	\$45.13	\$857.47 M
Semiconductor ETF	XSD	3/31/2006	0.35%	\$115.13	\$518.10 M
Metals & Mining ETF	XME	6/30/2006	0.35%	\$23.39	\$457.38 M
Oil & Gas Equipment & Services ETF	XES	6/30/2006	0.35%	\$43.91	\$121.41 M
Oil & Gas Exploration & Production ETF	XOP	6/30/2006	0.35%	\$66.61	\$2,337.89 M
Pharmaceuticals ETF	XPH	6/30/2006	0.35%	\$43.79	\$249.63 M
Regional Banking ETF	KRE	6/30/2006	0.35%	\$44.87	\$1,503.21 M
Retail ETF	XRT	6/30/2006	0.35%	\$44.21	\$362.53 M
Health Care Equipment ETF	XHE	3/31/2011	0.35%	\$88.00	\$521.40 M
Telecom ETF	XTL	3/31/2011	0.35%	\$73.55	\$53.32 M
Transportation ETF	XTN	3/31/2011	0.35%	\$57.37	\$174.97 M
Aerospace & Defense ETF	XAR	9/30/2011	0.35%	\$98.20	\$1,571.20 M
Health Care Services ETF	XHS	9/30/2011	0.35%	\$71.93	\$90.64 M
Software & Services ETF	XSW	9/30/2011	0.35%	\$110.64	\$236.77 M
The Real Estate Select Sector Fund	XLRE	12/31/2015	0.13%	\$37.43	\$4,641.12 M
Internet ETF	XWEB	6/30/2016	0.35%	\$98.83	\$21.74 M
The Communication Services Select Sector Fund	XLC	6/29/2018	0.13%	\$56.41	\$9,803.64 M

**Table 18**List of Sector SPDR ETFs. The expense ratio, Net Asset Value per share (NAV) and Assets Under Management (AUM) are as of 6/9/2020.



	Volume	Drift	Volatility
$Treated \times Post$	-0.0776 (0.435)	-0.00715** (0.003)	0.0128* (0.008)
Model	-0.0881	-0.00006	0.0112
In ETFs	343	343	343
In ETF Sectors	3316	3316	3316
Outside ETF Sectors	866	866	866
Time FE	YES	YES	YES

**Table 19 Effect of Introducing Sector ETFs.** Coefficients from:  $Outcome_{i,t} = \alpha + \beta \times Treated_{i,t} \times Post_t + \gamma_t + \epsilon_{i,t}$  Observations are weighted by lagged market capitalization. Standard errors, clustered at the firm level, in parenthesis.

	Baseline	Placebo		Baseline	Placebo	
		t=-22	FOMC		t=-22	FOMC
Inc. Passive	0.106*** (0.036)	-0.00474 (0.013)	0.00709 (0.009)	0.382** (0.180)	-0.0501 (0.042)	0.0159 (0.023)
Observations	126,319	157,769	126,654	126,319	157,769	126,654
R-squared	0.03	0.031	0.03	0.035	0.035	0.034
Controls/FE	Yes	Yes	Yes	Yes	Yes	Yes
Weights	Equal	Equal	Equal	Value	Value	Value

**Table 20 Placebo Test: Earnings Day Share of Volatility.** Table with estimates of  $\beta$  from:

$$\Delta QVS_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , 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  $t$ . Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t-1$  to  $t$ . Fixed effects are year and firm. Standard errors are computed using panel Newey-West with 2 lags. The “Baseline” columns use actual earnings dates, while the “Placebo” results are the coefficient estimates when selecting dates between the actual earnings days  $t = -22$ , or the FOMC meeting dates.

	(1)	(2)	(3)
Inc. Passive	-12.26*** (1.923)	-13.97*** (2.214)	-15.91** (6.207)
Observations	151,068	151,064	151,064
R-squared	0.03	0.031	0.181
Controls	No	Yes	Yes
Firm FE	No	No	Yes
Weight	Eq.	Eq.	Val.

**Table 21 Post-2000: Pre-Earnings Volume.** Estimates of  $\beta$  from:

$$\Delta CAV_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$  is cumulative abnormal pre-earnings trading volume.  $\Delta$  is a year-over-year change, matching on fiscal quarter. Change in passive ownership is expressed as a decimal, so  $0.01 = 1\%$  increase. Controls,  $X_{i,t-1}$ , include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, total institutional ownership and growth in market capitalization from  $t - 1$  to  $t$ . Standard errors are computed using panel Newey-West with 8 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter.

	(1)	(2)	(3)
Inc. Passive	-0.0282** (0.012)	-0.0273** (0.014)	-0.0660** (0.029)
Observations	151,023	151,020	151,020
R-squared	0.022	0.023	0.072
Controls	No	Yes	Yes
Firm FE	No	No	Yes
Weight	Eq.	Eq.	Val.

**Table 22 Post-2000: Pre-Earnings Drift.** Table with estimates of  $\beta$  from:

$$\Delta DM_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

Where  $DM_{i,t}$  is a measure of the pre-earnings drift. Passive ownership is expressed as a decimal, so  $0.01 = 1\%$  of shares outstanding held by passive funds. Controls,  $X_{i,t-1}$ , include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, total institutional ownership and growth in market capitalization from  $t - 1$  to  $t$ . Standard errors are computed using panel Newey-West with 8 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter.

	(1)	(2)	(3)
Inc. Passive	0.216*** (0.032)	0.140*** (0.035)	0.365** (0.172)
Observations	68,142	68,126	67,224
R-squared	0.012	0.013	0.037
Controls	No	Yes	Yes
Firm FE	No	No	Yes
Weight	Eq.	Eq.	Val.

**Table 23 Post-2000: Earnings Day Share of Volatility.** Table with estimates of  $\beta$  from:

$$\Delta QVS_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , 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  $t$ . Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t - 1$  to  $t$ . Fixed effects are year and firm. Standard errors are computed using panel Newey-West with 2 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-year lagged market capitalization relative to other firms that year.

		Baseline		AT Measures	
	(1)	(2)	(3)	(4)	(5)
Ch. Passive	-7.602*** (2.313)	-7.261** (3.385)	-19.28*** (6.339)	-1.651 (3.289)	-13.59** (6.338)
Ch. Oddlot				-3.428*** (0.272)	-2.596*** (0.712)
Ch. Trade/Order				6.009*** (0.245)	3.371*** (0.479)
Ch. Trade/Cancel				-1.941*** (0.245)	-4.248*** (0.595)
Ch. Tradesize				1.247*** (0.452)	2.789*** (1.015)
Obs	44,544	44,542	44,542	44,542	44,542
R-squared	0.053	0.078	0.171	0.19	0.262
Controls	No	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	Yes	Yes
Weight	Eq.	Eq.	Vw.	Eq.	Vw.

**Table 24 AT Activity: Pre-Earnings Volume.** Estimates of  $\beta$  from:

$$\Delta CAV_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$CAV_{i,t}$  is cumulative abnormal pre-earnings trading volume.  $\Delta$  is a year-over-year change, matching on fiscal quarter. Change in passive ownership is expressed as a decimal, so  $0.01 = 1\%$  increase. Controls,  $X_{i,t-1}$ , include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, total institutional ownership and growth in market capitalization from  $t - 1$  to  $t$ . Standard errors are computed using panel Newey-West with 8 lags. Only uses data that can be matched to MIDAS from 2012-2018. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter. Columns 4 and 5 add the AT activity measures from Weller (2017).

	Baseline			AT Measures	
	(1)	(2)	(3)	(4)	(5)
Ch. Passive	-0.0531*** (0.017)	-0.0770*** (0.024)	-0.0994*** (0.037)	-0.0790*** (0.024)	-0.103*** (0.037)
Ch. Oddlot				0.00408*** (0.001)	0.00158 (0.003)
Ch. Trade/Order				0.0012 (0.001)	0.00128 (0.003)
Ch. Trade/Cancel				0.00328*** (0.001)	0.00781*** (0.003)
Ch. Tradesize				0.00202 (0.002)	0.00551 (0.005)
Obs	44,527	44,525	44,525	44,525	44,525
R-squared	0.013	0.059	0.05	0.06	0.053
Controls	No	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	Yes	Yes
Weight	Eq.	Eq.	Vw.	Eq.	Vw.

**Table 25 AT Activity: Pre-Earnings Drift.** Table with estimates of  $\beta$  from:

$$\Delta DM_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

Where  $DM_{i,t}$  is a measure of the pre-earnings drift. Passive ownership is expressed as a decimal, so  $0.01 = 1\%$  of shares outstanding held by passive funds. Controls,  $X_{i,t-1}$ , include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, total institutional ownership and growth in market capitalization from  $t - 1$  to  $t$ . Standard errors are computed using panel Newey-West with 8 lags. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter.

	Baseline		AT Measures		
	(1)	(2)	(3)	(4)	(5)
Ch. Passive	0.312*** (0.050)	0.255*** (0.077)	0.789** (0.328)	0.237*** (0.077)	0.805** (0.334)
Ch. Oddlot				-0.00439 (0.005)	-0.00685 (0.030)
Ch. Trade/Order				-0.00875 (0.006)	-0.0384 (0.043)
Ch. Trade/Cancel				-0.0106** (0.005)	-0.0653 (0.053)
Ch. Tradesize				-0.0223** (0.009)	0.0211 (0.067)
Obs	19,220	18,815	18,815	18,815	18,815
R-squared	0.006	0.065	0.084	0.066	0.087
Controls	No	Yes	Yes	Yes	Yes
Firm FE	No	Yes	Yes	Yes	Yes
Weight	Eq.	Eq.	Vw.	Eq.	Vw.

**Table 26 AT Activity: Earnings Day Share of Volatility.** Table with estimates of  $\beta$  from:

$$\Delta QVS_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$$

$QVS_{i,t} = \sum_{\tau=1}^4 r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$ , 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  $t$ . Controls in  $X_{i,t-1}$  include 1-year lagged passive ownership, market capitalization, idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. I also condition on the growth in market capitalization from  $t-1$  to  $t$ . Fixed effects are year and firm. Standard errors are computed using panel Newey-West with 2 lags. Only uses data that can be matched to MIDAS from 2012-2018. Column 1 is a univariate regression, while column 2 includes the controls, as well as year/quarter and firm fixed effects. Column 3 is the same as column 2, except firms are weighted by their 1-quarter lagged market capitalization relative to other firms that quarter. Columns 4 and 5 add the AT activity measures from Weller (2017).



	AT Activity Score		1-Year Ch. In Score		3-Year Ch. In Score	
Level of Passive Ownership	1.585*** (0.155)	0.318** (0.158)				
1-Year Inc. in Passive			0.736*** (0.199)	0.369* (0.212)		
3-Year Inc. in Passive					0.619*** (0.211)	0.462 (0.290)
Observations	17,210	17,210	17,068	17,068	12,783	12,783
R-squared	0.3	0.086	0.093	0.098	0.125	0.108
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 27 Relationship Between Passive Ownership and AT Activity** I calculate the AT activity score as the first principal component of the 4 AT measures in Weller (2017). This score is normalized to have mean zero and standard deviation one. The effect in levels of moving from the 25th percentile (0.0) to 75th percentile (0.1) of passive ownership is a 0.15 standard deviation increase in AT activity score. Although a 0.15 standard deviation increase may seem small, this is about half the mean for firms with high passive ownership. Controls include firm market capitalization, institutional ownership, idiosyncratic volatility and market cap. growth.

	(1)	(2)	(3)	(4)
SUE	0.00912*** (0.000)		0.00314*** (0.000)	
SUE > 0		0.00745*** (0.000)		0.00369*** (0.000)
SUE < 0		-0.00394*** (0.000)		0.000128 (0.001)
SUE x Passive	0.0545*** (0.003)		0.0435*** (0.007)	
SUE > 0 x Passive		0.0217*** (0.003)		0.0246*** (0.006)
SUE < 0 x Passive		-0.0411*** (0.004)		-0.0196* (0.011)
Observations	415,961	415,961	415,961	415,961
R-squared	0.068	0.069	0.039	0.041
Controls	Yes	Yes	Yes	Yes
Firm FE	Yes	Yes	Yes	Yes
Weight	Eq.	Eq.	Val.	Val.

**Table 28 Passive Ownership and Response to Earnings News.** This table contains the results of the following regression:

$$r_{i,t} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 (SUE_{i,t} \times Passive_{i,t}) + \xi X_{i,t} + \text{Fixed Effects} + e_{i,t}$$

Here,  $r_{i,t}$  denotes the market-adjusted return on the effective quarterly earnings date.  $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$ . Controls in  $X_{i,t}$  include 1-year lagged passive ownership, market capitalization, growth of market capitalization from  $t - 1$  to  $t$ , idiosyncratic volatility, calculated as the sum of squared market-adjusted returns over the past year, and total institutional ownership. Fixed effects include year/quarter and firm. Columns 1 and 2 are equal weighted, while Columns 3 and 4 are value weighted.

		1-year		IVD 3-year		5-year	
Increase in Passive Ownership	1-year	0.325 (0.225)	0.334* (0.191)				
	3-year			0.173 (0.154)	0.100 (0.158)		
	5-year					0.328** (0.141)	0.300* (0.157)
Observations		4,519	4,519	3,979	3,979	3,496	3,496
		1-year		Term 3-year		5-year	
Increase in Passive Ownership	1-year	9.375 (5.873)	9.155** (4.495)				
	3-year			8.789* (4.758)	8.278** (4.000)		
	5-year					5.489 (4.547)	7.653* (4.121)
Observations		4,457	4,457	3,916	3,916	3,441	3,441
Firm Controls		Yes	Yes	Yes	Yes	Yes	Yes
Year FE		Yes	Yes	Yes	Yes	Yes	Yes
Industry FE		Yes	Yes	Yes	Yes	Yes	Yes
Firm FE		No	Yes	No	Yes	No	Yes

**Table 29 Option Regressions.** Estimates of  $\beta$  from:

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

Where the outcomes are  $IVD = \frac{\overline{IV}_b}{0.5(\overline{IV}_a + \overline{IV}_c)}$  and  $\text{Term} = \frac{(\sigma_{term}^Q)^2}{IV_b}$ . Sample includes 306 S&P 500 firms with options that meet the filters described in Kelly et al. (2016), and have at least 16 years with 4 non-missing earnings announcements. Controls in  $X_{i,t}$  include institutional ownership, lagged institutional ownership, market capitalization, lagged market capitalization. Fixed effects include industry, year and firm.