Passive Ownership and Price Informativeness

MARCO SAMMON*

Job Market Paper

December 14, 2020

ABSTRACT

Despite the rapid growth of passive ownership over the past 30 years, there is no consensus on how or why passive ownership affects stock price informativeness. This paper provides a new answer to this question by examining how passive ownership changes investors' incentives to acquire information. I develop a model where passive ownership affects how many investors gather information and how investors allocate attention between systematic and idiosyncratic risk. The model also links investors' learning decisions to price informativeness through quantities that are readily observable in the data: trading volume, returns and volatility. The model's predictions motivate three new measures of price informativeness, all of which declined on average over the past 30 years. In the cross-section, increases in passive ownership are negatively correlated with price informativeness. To establish causality, I show that price informativeness decreases after quasi-exogenous increases in passive ownership arising from index additions and rebalancing.

JEL classification: G12, G14.

1   Introduction

The rise of passive ownership is one of the most significant 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 owing nearly 15% in 2018. As passive ownership continues to grow, academics and practitioners are asking: 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) provide evidence that passive ownership increases non-fundamental volatility and makes prices less informative[1]. Given that passive funds trade on mechanical rules, the standard intuition is that as more investors become passive, there will be 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 with the best earnings news. In the early 1990s, 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. This is evidence of fundamental information being traded into prices before it is formally announced (see e.g., Ball and Brown (1968), Fama et al. (1969)).

Compare these patterns to what we see after 2010: The pre-earnings drift upward is smaller, and the return 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 the information is made public. From this comparison, it appears that pre-announcement prices were more informative 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

---

[1]Buss and Sundaresan (2020) also provide evidence that passive ownership makes prices more informative. Israeli et al. (2017), Bhattacharya and O'Hara (2018), Malikov (2018), Garleanu and Pedersen (2018), Chinco and Fos (2019) and Breugem and Buss (2019) provide evidence that passive ownership makes prices less informative.
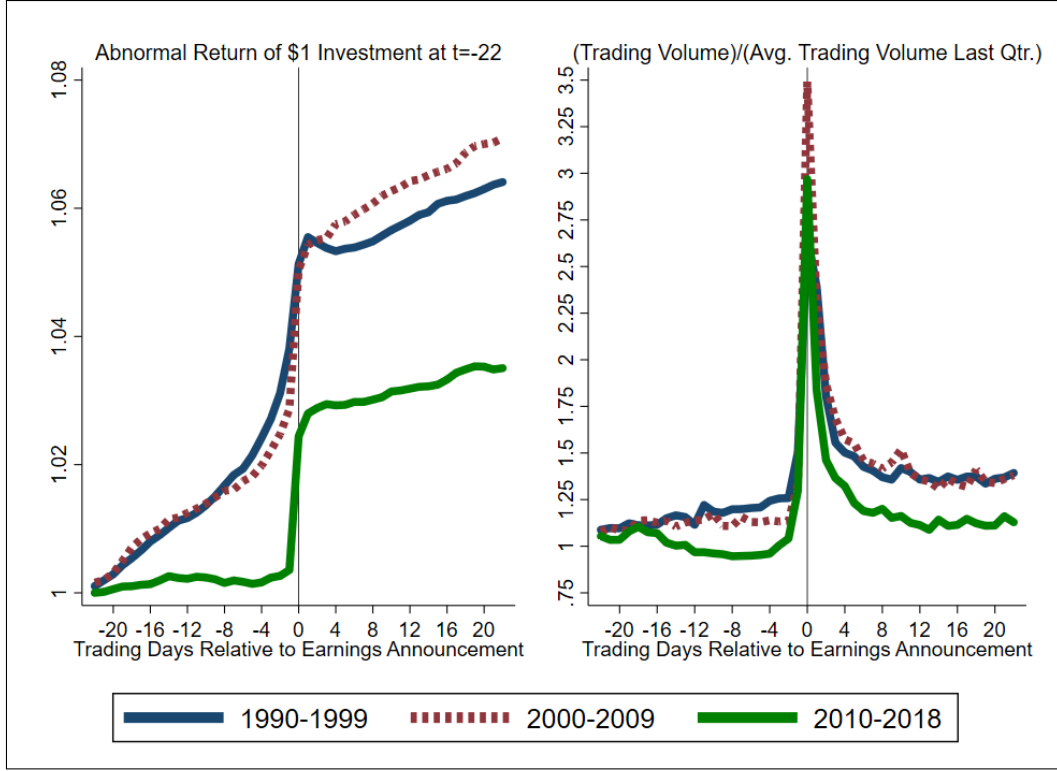
Figure 1. Returns and Trading Volume around Earnings Announcements. Each plot represents the cross-sectional average for firms in the top decile of Standardized Unexpected Earnings (SUE). SUE is defined as: $SUE_{i,t} = \frac{Earnings_{i,t} - Earnings_{i,t-4}}{\sigma_{(t-1,t-8)}(Earnings_{i,t} - Earnings_{i,t-4})}$. 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.

after the 2000s, it is unlikely that this trend was driven entirely by changes in disclosure rules (e.g., Regulation Fair Disclosure, passed in August 2000), and changes in the enforcement of insider trading laws (Coffee Jr (2007)).

Two trends are not causal. To formalize the relationship between passive ownership and price informativeness, a model is needed to (1) guide the measurement of price informativeness and (2) understand the mechanism. A natural starting point is Grossman and Stiglitz (1980): Because passive funds trade based on mechanical rules, they seem like uninformed investors. As the share of uninformed investors increases, price informativeness decreases. The Grossman-Stiglitz model, however, does not necessarily apply to the rise of passive ownership for three reasons.

3

First, passive ownership is not necessarily uninformed. According to the former head of BlackRock's ETF business, *"Only about 30 percent of ETF investors look at these as passive funds, who are just there long term.*[2]*"* Passive *funds* are heavily traded by active *investors*.

Second, passive ownership may affect not only who becomes informed, but also which risks investors learn about. For example, many ETFs are held by sophisticated institutional investors looking for targeted exposure to systematic risk factors. In reference to Global X's ETF offerings, its former CEO 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.*[3]*"* It's possible that the availability of ETFs leads investors to learn about systematic risks, rather than stock-specific risks.

Finally, Grossman-Stiglitz is hard to bring to the data. In the model, price informativeness is the conditional variance of fundamentals, given prices. The empirical analogue is a regression of future fundamentals on current prices:

$$\text{fundamentals}_{i,t+1} = \hat{\alpha} + \hat{\beta} \times \text{price}_{i,t} + \text{controls} + \hat{\epsilon}_{i,t} \tag{1}$$

Larger values of $\hat{\beta}$ suggest that prices are more informative: Fundamentals and prices covary more strongly with one another.

There are several empirical challenges when estimating this regression. First, the correct measure of future fundamentals is not obvious. In a static model like Grossman-Stiglitz, there are no cashflows after $t + 1$, but in reality, firms are long-lived. Maybe the left-hand-side variable should be *all* fundamentals from $t+1$ forward, which are hard to measure. The right set of conditioning variables is also not clear: An econometrician does not know which variables investors use along with the price when forming expectations.

In this paper, I develop a model to address these issues. Passive ownership is 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.

---

[2]Daniel Gamba, Global Head of Active Equity Product Strategy, BlackRock, quoted in Balchunas (2016).
[3]Bruno del Ama, former CEO of Global X, quoted in Balchunas (2016).

The model also guides a new way of linking observable quantities to investors' learning behavior and price informativeness.

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 investors and increases attention 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 access to a well-diversified portfolio. This tends to increase the share of uninformed investors.

Which channels dominate depend on model parameters. The natural way to resolve this ambiguity is to calibrate the model to match the data. The share of informed investors, and which risks investors learn about, however, is not readily observable. Instead, I focus on the model's predictions for quantities easily measured in the data: returns, volume and volatility. I use these predictions to define three new 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 measures of price informativeness 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 announcements. Passive ownership is also correlated with decreased pre-earnings drift and increased volatility on earnings days. To rule out the possibility that these results are driven by simultaneous trends, or regime shifts in financial markets (e.g., changes in disclosure rules), all the cross-sectional regressions include year/quarter fixed-effects and are run in first differences. These cross-sectional results imply that passive ownership decreases price informativeness.

The model can qualitatively match the cross-sectional regression results when calibrated to quantitatively match the observed rise in passive ownership. I provide direct evidence of

5

the model's predictions for the effect of passive ownership on investors' learning behavior: Increases in passive ownership are correlated with decreased analyst coverage, less accurate analyst forecasts, more time between analyst updates and decreased downloads of SEC filings. This, however, raises the concern of reverse causality: Perhaps passive ownership increased 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 increases in passive ownership that are plausibly uncorrelated with firm fundamentals. To this end, I design two natural experiments using 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 announcement price informativeness between 1990 and 2018. Section 4 links the trends in passive ownership and price informativeness 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 after these quasi-exogenous increases in passive ownership.

2   Model of learning and passive ownership

In this section, I incorporate passive ownership into an Admati (1985)-style 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 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.

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) 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. I restrict to equilibria where there are a positive measure of informed investors.

Assets

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

$$z_i = a_i + f + \eta_i \tag{2}$$

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

In the baseline version of the model, stocks are symmetric: $\sigma_i^2 = \sigma^2$, $\overline{x}_i = \overline{x}$ and, $\sigma_{i,x}^2 = \sigma_x^2$. This assumption is not needed, but it simplifies the intuition for the key learning

---

[4]There exists an equivalent economy where stock returns have the same correlation structure, but there is no systematic risk-factor. For example, suppose $cov(\eta_i, \eta_j) = \sigma_f^2$ 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. The Online Appendix discusses this representation issue in more detail.

trade-offs. For an extension where individual stocks load differently on systematic risk, and have heterogeneous volatility of their idiosyncratic risk-factors, see the Online Appendix.

I also assume that the number of stocks n is sufficiently small so that idiosyncratic risk cannot be totally diversified away. Explicitly, $Var(\frac{1}{n}\sum_{i=1}^{n} z_i) > Var(f)$. 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 an investor wanted to trade an 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*:

$$s_{i,j} = a_i + (f + \epsilon_{f,j}) + (\eta_i + \epsilon_{i,j}) \tag{3}$$

where $\epsilon_{i,j} \overset{iid}{\sim} N(0, \sigma^2_{\epsilon_{i,j}})$, $\epsilon_{f,j} \sim N(0, \sigma^2_{\epsilon_{f,j}})$ and $\epsilon_{i,j}$ are independent for all permutations of i and j, as well as independent from $\epsilon_{f,j}$. The signal noise, ϵ, depends on how much attention investor j devotes to each *risk-factor* that affects the payoff of stock i: η_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 η_i or f to reduce signal noise:

$$\sigma^2_{\epsilon_{i,j}} = \frac{1}{\alpha + K_{i,j}}, \quad \sigma^2_{\epsilon_{f,j}} = \frac{1}{\alpha + K_{n+1,j}} \tag{4}$$

where $\alpha > 0$. This differs from Kacperczyk et al. (2016), where the learning technology is $\sigma^2_{\epsilon_{i,j}} = \frac{1}{K_{i,j}}$. In my setting, $\sigma^2_{\epsilon_{i,j}}$ needs to be well defined even if an investor devotes no attention to risk-factor η_i or f^6. α 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 model's

---

[5]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.

[6]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 assets 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.

predictions to α in the Online Appendix.

Informed investors have a total attention constraint of $\sum_{i,j} K_{i,j} \leq K$. They also have a no forgetting constraint, so $K_i \geq 0$ for all i. In the baseline calibration, learning capacity K is fixed to 1. The Online Appendix discusses an alternative version of the model where investors can pay to increase learning capacity.

Portfolio Choice

Define terminal wealth:

$$w_{2,j} = (w_{0,j} - \mathbb{1}_{informed,j}c) + \mathbf{q}'_j(\mathbf{z} - \mathbf{p}) \tag{5}$$

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:

$$U_{1,j} = E_{1,j}[-exp(-\rho w_{2,j})] \tag{6}$$

where ρ 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. The assumption of no independent signal about the $(n+1)^{th}$ risk-factor is needed to solve the model using the closed form solutions in Admati

$(1985)^7$.

To solve for prices, start by defining μ as the vector of a_i's. Further define $\overline{\mathbf{x}}$ as the vector of $\overline{x}_i$'s. Define the $n \times (n+1)$ matrix Γ as $[I_n \quad 1_{n,1}]$ i.e. concatenating an $n \times n$ identity matrix with a $n \times 1$ vector of 1's.

Defining η as a vector of η_i's and f (where f is the last entry), terminal asset payoffs are $\mathbf{z} = \mu + \Gamma\eta$. Define the variance of stock payoffs as the matrix V, and the matrix of stock signal variances for investor j as S_j. I assume all informed investors have the same attention allocation, so $S_j = S$ for all j and $K_{i,j} = K_i^8$. Given the learning technology, S_j^{-1} will always be positive definite for informed investors

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 $\overline{\mathbf{x}} + \mathbf{x}$ 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):

$$\mathbf{p} = A_0 + A_1\mathbf{z} - A_2(\overline{\mathbf{x}} + \mathbf{x}) \tag{7}$$

Beliefs

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

$$s_p = A_1^{-1}\left(p - A_0 + A_2(\overline{\mathbf{x}} + \mathbf{x})\right) \tag{8}$$

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.

---

[7]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 $p = \tilde{A}_0 + \tilde{A}_1\eta + \tilde{A}_2 f + \tilde{A}_3\mathbf{x}$, where η 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.

[8]For discussions of non-symmetric equilibria, see e.g., Veldkamp (2011).

Demands

Demands are a function of private signals and prices. There are separate demand functions for the informed and uninformed:

$$\text{Uninformed: Demand} = G_0 + G_{2,un}\mathbf{p}$$
$$\text{Informed, investor } j\text{: Demand} = G_0 + G_1\mathbf{s_j} + G_{2,inf}\mathbf{p} \tag{9}$$

where $\mathbf{s}_j$ is the vector of signals received by investor j[9].

Deciding to Become Informed

I follow Kacperczyk et al. (2016) and give investors a preference for the early resolution of uncertainty. 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: $U_{0,j} = -E_0[ln(-U_{1,j})]/\rho$ where the time 0 information set is the share of investors who decide to become informed. This simplifies to:

$$U_{0,j} = E_0\left[E_{1,j}[w_{2,j}] - 0.5\rho Var_{1,j}[w_{2,j}]\right] \tag{10}$$

because time two wealth is normally distributed. The Online Appendix discusses how Equation 10 is derived from Epstein and Zin (1989) preferences, and how these preferences 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:

$$z_{n+1} = a_{n+1} + f \tag{11}$$

When it is present, the ETF initially has average supply $\bar{x} = 0$, but is still subject to normally-distributed supply shocks x_{n+1}. 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. 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

---

[9]Details on the demand functions are in Appendix A.1.

(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[10].

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[11].

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

$$U_{0,j} = E_0\left[E_{1,j}[w_{2,int}] - 0.5\rho^{int}Var_{1,j}[w_{2,int}]\right] \tag{12}$$

where ρ^{int} 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 $v/\overline{x_i}$ 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:

$$w_{2,int} = v\left(\underbrace{\sum_{i=1}^{n}(z_i - p_i)}_{\text{Idio. Risk}} - \underbrace{n(z_{n+1} - p_{n+1})}_{\text{Sys. Risk}}\right) \tag{13}$$

---

[10]The Online Appendix presents a micro-foundation for the strcture off supply shocks to the ETF.

[11]The Online Appendix contains a more thorough discussion of the implications of these two assumptions.

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.

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 ρ^{int}: 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 would be consistent with a decrease in ρ^{int}. 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 ρ, σ, σ_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. The Online Appendix discusses how sensitive passive ownership is to these parameter choices.

To summarize, there are two ways to increase the size of passive ownership through agents' preferences: (1) Decrease the intermediary's risk aversion ρ^{int} or (2) Increase informed and uninformed investors' risk aversion ρ. Having these two channels in the model, instead of just directly varying $v/\overline{x}$, is important for thinking about causality in the empirical results. A decrease in ρ^{int} leading increased v and changes in learning implies causality coming from passive ownership. An increase in ρ, however, would imply the reverse: the intermediary creates more shares of the ETF *because* of changes in investors' demand functions, which itself is a function of changes in learning behavior.

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:

$$s_{i,j} = a_i + (f + \epsilon_{f,j}) + (\eta_i + \epsilon_{i,j}) \text{ for } i = 1, \ldots, n$$
$$s_{n+1,j} = a_i + (f + \epsilon_{f,j})$$

(14)

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 $\overline{x} - v$ 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

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)[12].

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. In addition, ETFs cover more sectors/indexes than futures contracts and mutual funds.

---

[12]Quoted in Balchunas (2016)

2.2   Equilibrium and Learning Trade-Offs

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[13].

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 (σ_f^2) is low, relative to idiosyncratic risk-factors (σ^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[14]. The Online Appendix presents two asset examples of these learning trade-offs.

2.3   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.[15]

2.3.1   Intensive Learning Margin

Changing the share of informed investors affects which risks investors learn about in equilibrium. To shut off this channel, and isolate the intensive margin effects of passive ownership, I fix the share of informed investors. I compare attention to systematic risk

---

[13]See Section A.3 of the Appendix for a step-by-step outline of the numerical solution method.

[14]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.

[15]The parameters are mostly taken from Kacperczyk et al. (2016). See Appendix A.2 and the Online Appendix for details

across three scenarios: (1) *No ETF*, this is when investors cannot trade the ETF (2) *High* ρ^{int}, 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* ρ^{int}, investors can trade the ETF and the intermediary is closer to risk neutral, so the ETF is in larger supply.

Figure 2 shows how attention to systematic risk changes as we vary the size of passive ownership, fixing the share of informed investors at 60%. The left panel examines the effect of varying risk aversion ρ, fixing the volatility of the systematic risk-factor σ_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.
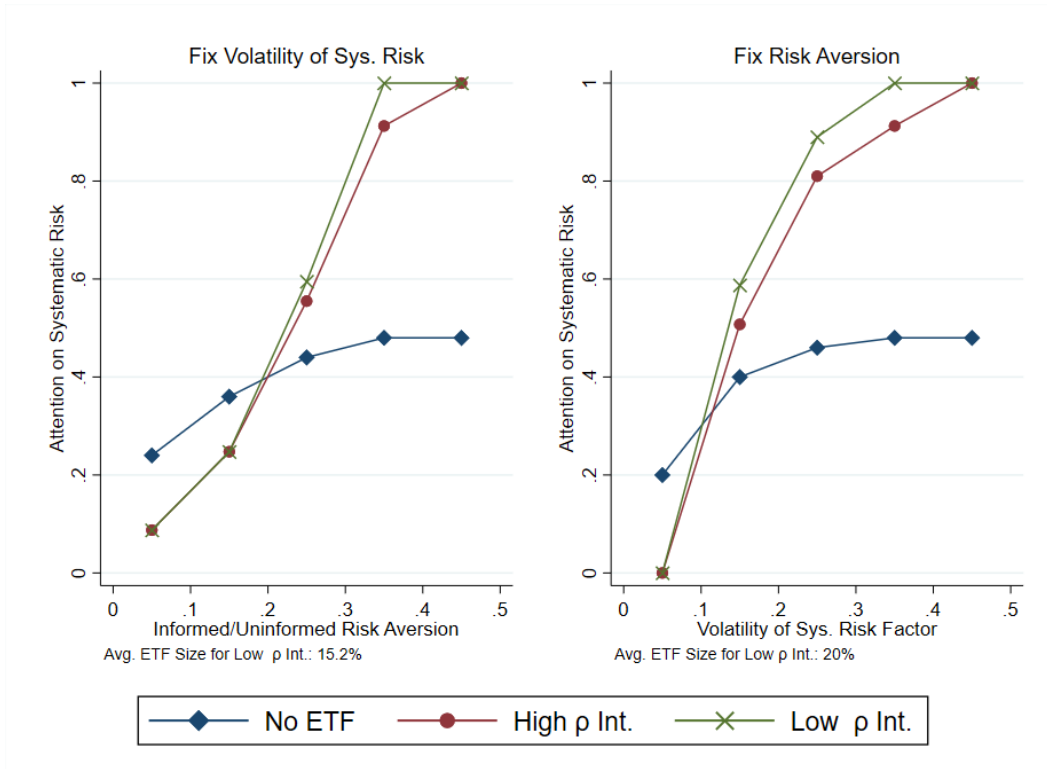

Figure 2. Effect of Introducing the ETF on Attention Allocation (fixed share informed). In both panels, the share of informed investors is fixed at 60%. In the left panel, $\sigma_f = 0.35$. In the right panel, $\rho = 0.35$. The y-axis reports the share of investors' attention devoted to the systematic risk-factor.

16

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 from Equation 9 is a measure of how sensitive demand is to their private signals. Table 1 contains selected the entries of G_1. As with Figure 2, 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[16].

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 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.

Figure 2 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

---

[16]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. See the Online Appendix for details.

			No ETF			ETF	
ρ	σ_f^2	$G_{i,i}$	$G_{i,j}$	$7 \times G_{i,j}$		$G_{i,i}$	$G_{i,j}$
0.1	0.2	0.968	-0.117	-0.817		1.260	-1.260
0.1	0.5	0.766	-0.069	-0.484		1.010	-1.010
0.35	0.2	0.189	-0.014	-0.100		0.046	-0.046
0.35	0.5	0.176	-0.012	-0.086		0.003	-0.003

Table 1 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" columns are the entries of G_1 after introducing the ETF in zero average supply. There are $n = 8$ stocks.

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 σ_f. The right panel of Figure 2 examines the effect of varying σ_f, fixing risk aversion ρ at 0.35. Increasing σ_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 σ_f is sufficiently low, increasing passive ownership can lead to less learning about systematic risk, while if σ_f is sufficiently high, the opposite is true.

2.3.2 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 scenarios: (1) no ETF (2) high ρ^{int} and (3) low ρ^{int}. Figure 3 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 investors, the volatility of the systematic risk factor and

informed/uninformed investors' risk aversion[17].

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.
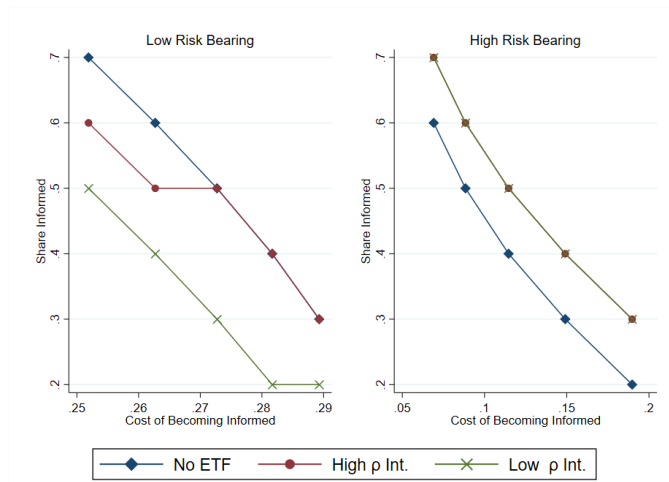

Figure 3. Effect of passive ownership on extensive learning margin. Left panel: $\sigma_f = 0.25$, $\rho = 0.15$. Right panel: $\sigma_f = 0.05$, $\rho = 0.05$. The x-axis is the cost in dollars of becoming informed. The y-axis reports the share of investors who become informed in equilibrium at this cost.

---

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

Figures 2 and 3 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 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.

2.4 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[18].

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.

---

[18]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. Further, 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)).

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 $\overline{x}$. 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.

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

$$\sum_{j}^{J} |\mathbf{q}_j - (\overline{\mathbf{x}} + \mathbf{x})/(J)| \tag{15}$$

where the first term $\mathbf{q}_j$ is investor j's demand, and the second term $(\overline{\mathbf{x}} + \mathbf{x})/(J)$ is investor j's share of the initial endowment $\overline{\mathbf{x}}$, 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 because they all use the same signal s_p from prices to form their posterior beliefs All investors have the same endowment, so if there were only uninformed investors, there would be no trading volume (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:

$$DM = \begin{cases} \frac{1 + r_{(0,1)}}{1 + r_{(0,2)}} & \text{if } r_2 > 0 \\ \frac{1 + r_{(0,2)}}{1 + r_{(0,1)}} & \text{if } r_2 < 0 \end{cases} \tag{16}$$

21

where $r_{(0,t)}$ is the cumulative market-adjusted return from 0 to t[19]. 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[20]. To compute this measure, I save the prices at $t = 0$, $t = 1$, $t = 2$, 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.

Share of Volatility on Earnings Days

Define the share of volatility on earnings days as:

$$\frac{r_2^2}{r_1^2 + r_2^2} \tag{17}$$

If prices are not informative before earnings announcements, we would expect earnings day volatility to be large, relative to total volatility[21].

Effect of Learning on Price Informativeness

It seems natural that these three measures of price informativeness should be related to the intensive and extensive learning margins. While the model does not offer closed-form expressions for these relationships, I examine them using simulated moments. To this end, I simulate the economy 10,000 times for particular choices of ρ, σ_f and the share of informed investors. Then, I calculate the average trading volume, drift and volatility across these simulations for assets 1 to n i.e. the stocks.

Ideally, I would directly vary informed investors' attention to the systematic risk factor. These choices, however, might not be optimal. Instead, I vary ρ as an indirect way of increasing attention on the systematic risk factor. Risk averse investors prefer to learn about systematic risk, relative to idiosyncratic risk, because volatile stock-specific risk factors can be diversified away. Shifting ρ is one way of varying attention on the systematic risk factor,

---

[19]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. The Online Appendix presents quantitative results on the relationship between passive ownership and risk premia.

[20]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.

[21]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., $|r_2| / (|r_1| + |r_2|)$.

while ensuring that attention is still optimally allocated.

Panel A of Figure 4 plots the attention on systematic risk and the three price informativeness measures against risk aversion, ρ. To isolate the intensive learning margin's effect on price informativeness, I fix the share of informed investors at 60%. As expected, increasing risk aversion leads to increased attention on the systematic risk factor. At the same time, increased risk aversion leads to decreased pre-earnings trading volume, decreased pre-earnings drift and increased earnings-day volatility. This is evidence that increases in attention to systematic risk lead to lower stock price informativeness.

Panel B plots the three measures of price informativeness against ρ for high and low shares of informed investors, highlighting the effect of the extensive learning margin. The blue line represents simulated model moments when the share of informed investors is low, at 20%, while the red line increases the share of informed investors to 70%. Across these two scenarios, I allow investors to optimally re-allocate attention. As expected, increasing the share of informed investors leads to increased pre-earnings trading volume, increased pre-earnings drift and decreased earnings-day volatility. This is evidence that increases in the share of informed investors lead to increased stock price informativeness.

Figure 4 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 from nothing to owning 15% of the US stock market. In Figure 5, I plot simulated model moments for two scenarios: (1) No ETF (2) ETF owing 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 investors when the ETF is not present. Then, I calculate how many investors optimally become informed in equilibrium at this cost when the ETF owns 15% of the market. To match the volatility of the S&P 500 in my empirical sample (1990-2018), I set $\sigma_f = 0.15$. All the price informativeness measures are only calculated for the stocks i.e. assets 1 to n.

The top 3 panels are averages of the price informativeness measures in an economy with low risk aversion $\rho = 0.05$. Increasing the share of informed investors (moving to the right along the x-axis) unambiguously increases price informativeness: it increases the pre-earnings
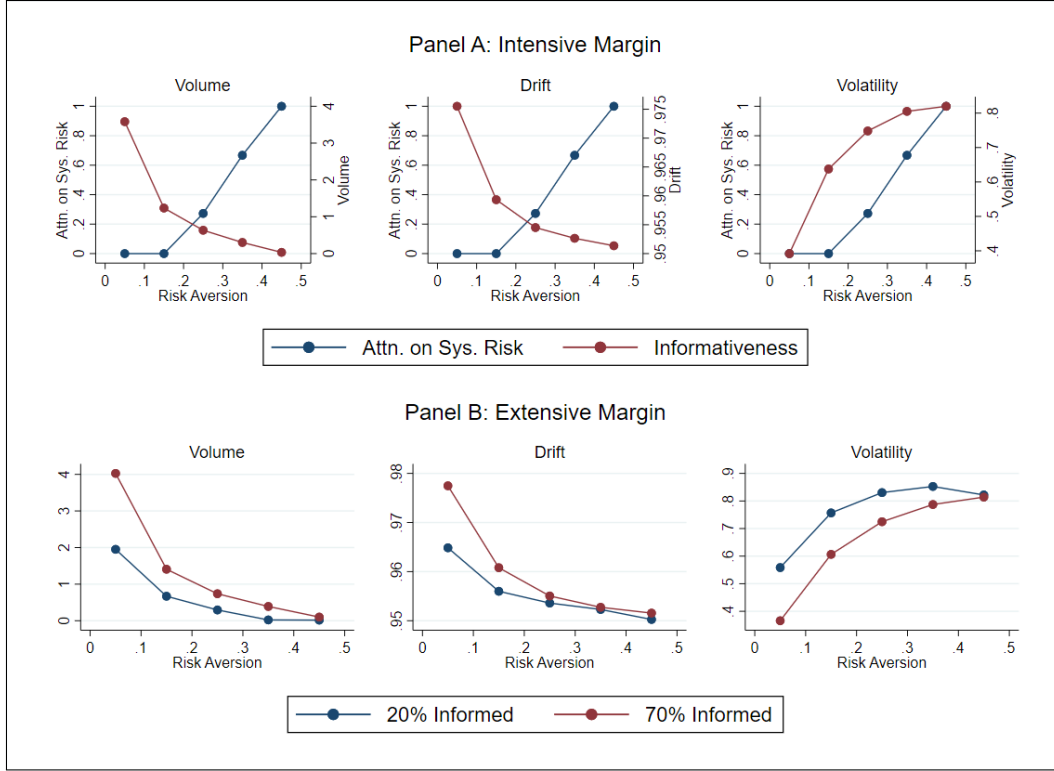
Figure 4.   Effect of Learning on Price Informativeness. Panel A plots the attention on systematic risk and three measures of price informativeness against informed and uninformed investors' risk aversion, ρ. Panel B plots the three measures of price informativeness against ρ for high and low shares of informed investors. In Panels A and B, the volatility of the systematic risk factor, $\sigma_f = 0.15$, and the risk aversion of the ETF intermediary, $\rho^{int}=1$. In Panel A the share of informed investors is set to 60%.

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 $\rho = 0.35$. 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 investors is larger because there is more room for the extensive
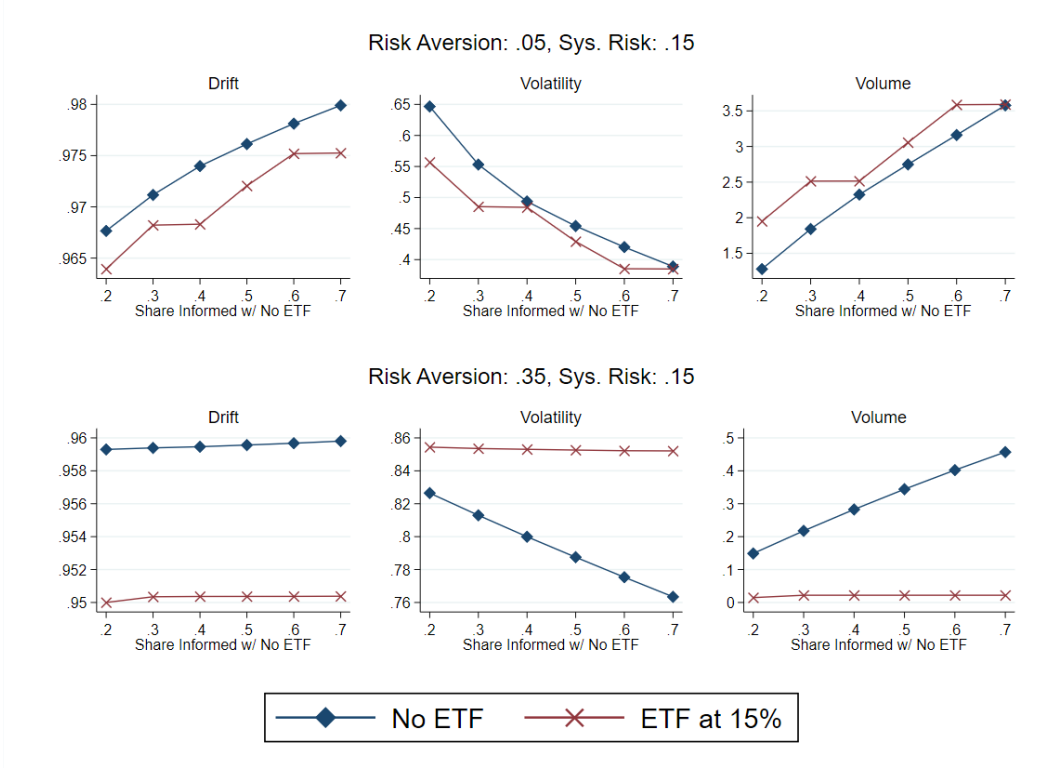
Figure 5. Effect of Passive Ownership on Price Informativeness. Top panels: $\rho = 0.05$. Bottom panels: $\rho = 0.35$. All panels: $\sigma_f = 0.15$. The cost of being informed is set so to match the share of investors who become informed when the ETF is not present.

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[22]. 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 $v/\overline{x}$, 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 6 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.

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

---

[22]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.
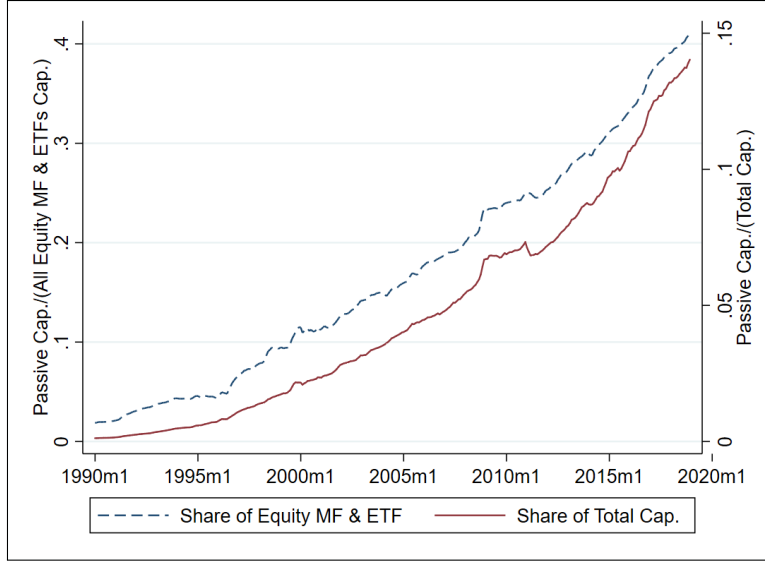
Figure 6. The Rise of Passive Ownership: 1990-2018. Passive funds are defined as all index funds, all ETFs, and all mutual funds with "index" in the name. Total equity mutual fund and ETF assets is the sum of all stock holdings in the Thompson S12 data that can be matched to CRSP.

effective earnings date (as long as the next day is not a trading holiday). If earnings are released Friday on or 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.[23] 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 investors are less willing to trade because

---

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

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 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:

$$AV_{i,t+\tau} = \frac{V_{i,t+\tau}}{\overline{V_{i,t-22}}} = \frac{V_{i,t+\tau}}{\sum\limits_{k=1}^{63} V_{i,t-22-k}/63} \tag{18}$$

Where abnormal volume, $AV_{i,t+\tau}$, is volume divided by the historical average volume for that firm over the past quarter[24] In Equation 18, $V_{i,t+\tau}$ 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:

$$AV_{i,t+\tau} = \alpha + \sum_{k=-21}^{22} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau} \tag{19}$$

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 ± 22 day window around each earnings

---

[24]All results are robust to instead using the average volume for that firm over the past year.

announcement.

I run this regression for three sample periods: (1) 1990-1999 (2) 2000-2009 (3) 2010-2018. Figure 7 plots the estimates of β_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.
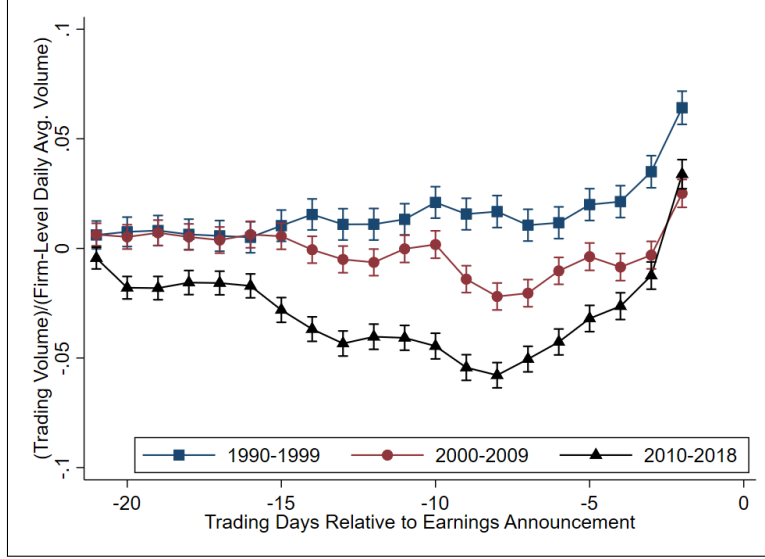

Figure 7. Decline of Pre-Earnings Volume. Plot of β_k estimated from the regression:

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

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

$$CAV_{i,t} = \sum_{\tau=-22}^{-1} AV_{i,t+\tau} \tag{20}$$

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. Between the 1990s and 2010s, value-weighted average $CAV_{i,t}$ declined by about 1. This can

be interpreted as a loss of 1 trading-day's worth of volume over the 22-day window before earnings announcements.

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[25]. 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, as in Kyle (1985).

Let $E_{i,t}$ denote earnings per share for firm i in quarter t in the IBES Unadjusted Detail File[26]. 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: $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$. 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 8 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[27]. 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

---

[25]See e.g.,, Ball and Brown (1968), Foster et al. (1984), Weller (2017).

[26]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

[27]For a detailed discussion of Reg FD, see the Online Appendix.

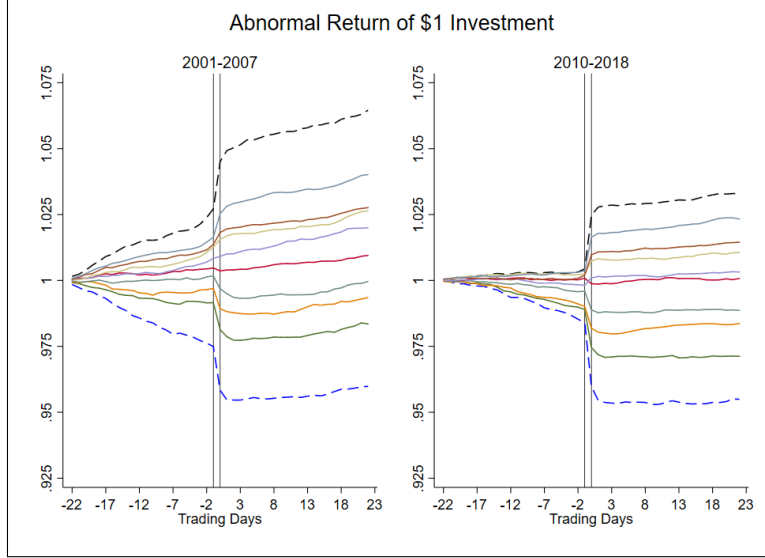in each decile move less before earnings days than between 2001 and 2007.

Figure 8. Decline of Pre-Earnings Drift by SUE Decile. Each quarter, I sort firms into deciles on standardized unexpected earnings. 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 8'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:

$$DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases} \tag{21}$$

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

31

would be reversed, which is why the measure is inverted when r_t is less than zero.[28] I work with gross returns, rather than net returns, to avoid dividing by numbers near zero[29]

$DM_{i,t}$ is 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. Value-weighted pre-earnings drift decreased by about -0.02 between 1990 and 2018.

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:

$$QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2 \tag{22}$$

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 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 9 plots coefficients from a regression of QVS on a set of year dummy variables for

---

[28]The Online Appendix presents alternative definitions of the pre-earnings drift using squared returns and further motivates my specification for $DM_{i,t}$.

[29]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.

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 12 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).
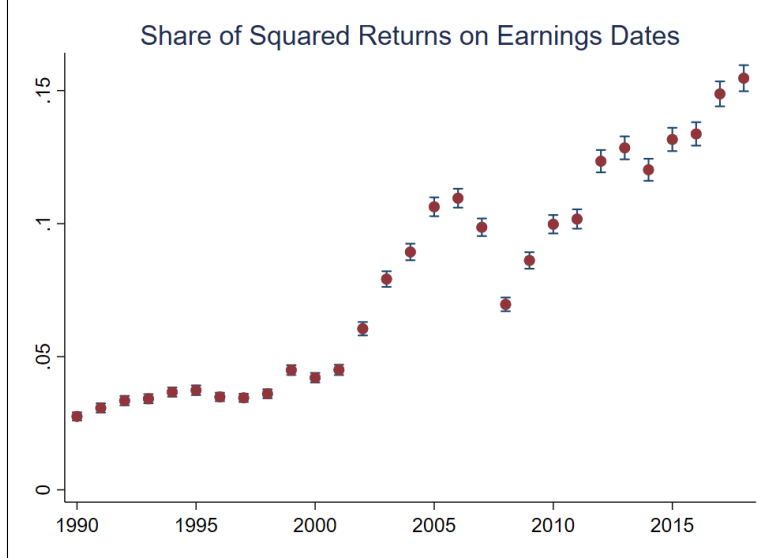

Figure 9. Increase in Earnings Day Volatility. This figure plots coefficients from a regression of QVS on a set of year dummy variables. For firm i in year t the quadratic variation share (QVS) is defined as: $QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, where r denotes a 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.

3.6   Robustness of Stylized Facts

These downward trends in price informativeness could be unrelated to the information released on earnings days. To rule this out, I run 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 the Online Appendix, I show that there is no drop in

volume before the placebo earnings dates. Further, there is no downward trend in the pre-earnings drift for the placebo earnings dates. Finally, there 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 placebo test, the Online 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. As shown in Section 2, 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 in pre-FOMC announcement volume, pre-FOMC announcement drift, or increased FOMC announcement dates' share of annual volatility at the stock-level. This confirms that the reduction in price informativeness only applies to firm-specific information.

4   Cross-sectional relationship between passive ownership and price informativeness

In this section, I show the cross-sectional 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 cross-sectional 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:

$$\Delta CAV_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \tag{23}$$

where cumulative abnormal pre-earnings volume, $CAV_{i,t}$, is defined in Equation 20. Δ is a year-over-year change, matching on fiscal quarter[30]. 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. These time fixed effects ensure I am comparing firms at the same point in time with different increases in passive ownership, conditional on their past level of passive ownership. This allays concerns that my results are driven by simultaneous trends in passive ownership and price informativeness. Although the time fixed effects account for *aggregate* trends in passive ownership and price informativeness, they do not account for *firm-specific* trends in either variable. This is also not solved by adding firm fixed-effects. Section B.2 of the Appendix shows that all of the baseline results are robust to running the regressions in levels with firm and time fixed effects. Standard errors are computed using panel Newey-West with 8 lags, and all results are robust to double-clustering standard errors at the firm/year level.

Table 2 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.

---

[30]One concern with the measure of abnormal volume in Equation 18 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.

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

Table 2 Passive Ownership and Pre-Earnings Volume. Estimates of β 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. 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. All specifications include year/quarter fixed effects. Standard errors are computed using panel Newey-West with 8 lags. Standard errors in parenthesis.

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:

$$\Delta DM_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \tag{24}$$

where $DM_{i,t}$ is defined as in Equation 21. 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 computing using panel Newey-West with 8 lags.

The regression results are in Table 3. The coefficient on $\Delta Passive_{i,t}$ in 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.

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

Table 3 Passive Ownership and Pre-Earnings Drift. Table with estimates of β 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. All specifications include year/quarter fixed effects. Standard errors are computed using panel Newey-West with 8 lags. Standard errors in parenthesis.

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:

$$\Delta QVS_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t} \tag{25}$$

where QVS is defined in Equation 22. 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 two lags[31]

The regression results are in Table 4. 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

---

[31]I use two lags, instead of eight as in the previous two sub-sections. This regression is run at the annual frequency, so there is only one overlapping observation, instead of four.

of 0.057 is about 1/2 of a standard deviation.

One explanation for the increased share of annual volatility on earnings dates is that the response to earnings news has increased. In the Online Appendix, I show this was the case between 1990 and 2018. The increase in response to earnings news, however, was especially strong for stocks with high passive ownership.

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

Table 4 Passive Ownership and Earnings Day Share of Volatility. Table with estimates of β 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. All specifications include year fixed effects. Standard errors are computed using panel Newey-West with 2 lags. Standard errors in parenthesis.

4.4 Robustness of Cross-Sectional Regression 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. The Online Appendix 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 nth 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 on the same calendar date every year. The first point also applies to the days between earnings announcements. All of the 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)). The Online Appendix shows that all my results are robust to only using data after Reg FD passed. My results are also robust to controlling for the AT measures in Weller (2017). 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 cross-sectional 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 on a grid of (1) the share of informed investors when the ETF is not present (2) risk aversion ρ (3) the volatility of the systematic risk factor σ_f (4) risk aversion of the ETF intermediary ρ^{int}.

The results are in Table 5. 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,

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, ρ^{int} is set to 1.

The row labeled data is the effect of a 15% increase in passive ownership based on value-weighted cross-sectional 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 off two by orders of magnitude. 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 relevant point of comparison.

		Volume	Drift	Volatility
	No ETF	0.9898	0.96388	0.70028
Model	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% Δ Passive	-3.594	-0.0145	0.05715

Table 5 Calibrating the Model to Match the Cross-Sectional 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 ρ^{int} is set to 1. Data row is effect of a 15% increase in passive ownership based on value-weighted cross-sectional estimates.*

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 ρ or volatility of the systematic risk factor σ_f 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:

$$\Delta Outcome_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{26}$$

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 26, 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:

$$\Delta DL_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t} + \text{Fixed Effects} + e_{i,t} \tag{27}$$

where Δ 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[32]. 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.

Table 6 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.

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

Table 6 Investor Attention and Passive Ownership. This table contains the estimates of β 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. Standard errors in parenthesis. All these regressions are equal-weighted.

The negative correlation between passive ownership and information gathering, raises the possibility that the cross-sectional results are driven by reverse causality. Maybe passive ownership happened to increase the most in stocks that had the biggest decline in passive

---

[32]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.

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

As outlined in Section 2, there are two ways to increase the size of passive ownership through agents' preferences in the model: (1) Decrease the intermediary's risk aversion ρ^{int} or (2) Increase informed and uninformed investors' risk aversion ρ. To confirm that passive ownership is the cause of the decrease in price informativeness, I aim to identify increases in passive ownership corresponding to decreases in ρ^{int}. 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 and investors' preferences. These allow me to causally link increases in passive ownership and decreases in pre-earnings price informativeness.

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 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.

To identify the causal effect of passive ownership on stock price informativeness, I use index addition as the treatment in a difference-in-differences regression:

$$\Delta Outcome_{i,t} = \alpha + \beta \times Treated_{i,t} + FE + \epsilon_{i,t} \tag{28}$$

Where $Outcome_{i,t}$ is the average pre-earnings volume, drift, or earnings days' share of annual volatility in the two years before or after index addition[33]. 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. I also include industry and month of index addition fixed effects. Including these fixed effects ensures I am only comparing treatment and control firms at the same point in time.

Because the increase in passive ownership associated with being added to the index varies over time, I also run a specification that allows for heterogeneous treatment intensity:

$$\Delta Outcome_{i,t} = \alpha + \beta \left(\Delta Passive_{i,t} \times Treated_{i,t} \right) + FE + \epsilon_{i,t} \tag{29}$$

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 10 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

---

[33]I take the average over the pre and post periods to avoid downward bias in the standard errors, as discussed in Bertrand et al. (2004).

changes in passive ownership, although the firms already in the index have a higher average level of passive ownership.
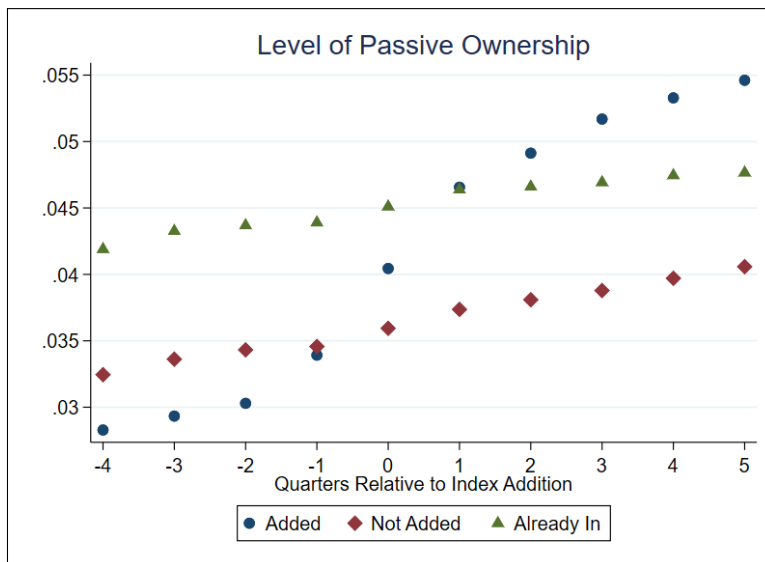

Figure 10. S&P 500 Index Addition and Changes in Passive Ownership. Average level of passive ownership for control firms out of the index ("Not Added"), control firms in the index ("Already In") and treated firms ("Added").

Table 7 contains the regression results. For comparison, I included a row with the reduced form estimates, which correspond to the value-weighted specification with all controls and fixed effects estimated in Section 4. For all three regressions, the results have the same sign as the cross-sectional regressions. The estimated coefficients, however, are larger than the cross-sectional results. I believe $\Delta Passive_{i,t}$ 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.

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. The Online Appendix has more details on the effect of S&P 500 index deletion.

	Binary Treatment		
	Volume	Drift	Volatility
Treated	-0.775**	-0.00583**	0.0186**
	(0.382)	(0.002)	(0.008)
R-squared	0.114	0.099	0.227
	Continuous Treatment		
Interaction	-27.15**	-0.158*	0.653**
	(12.680)	(0.096)	(0.281)
R-squared	0.114	0.096	0.227
Reduced Form	-23.96***	-0.0965***	0.381**

Table 7 Effects of S&P 500 Index Addition.
Estimates from:
$$\Delta Outcome_{i,t} = \alpha + \beta \times Treated_{i,t} + FE + \epsilon_{i,t}$$

And:
$$\Delta Outcome_{i,t} = \alpha + \beta \left(\Delta Passive_{i,t} \times Treated_{i,t} \right) + FE + \epsilon_{i,t}$$

Where $Treated_{i,t}$ is a dummy variable equal to one if a firm was added to the S&P 500 index. Both specifications include month of index addition and industry fixed effects. Standard errors in parenthesis. Treated firms: 419, Control Firms (Out of Index): 906, Control Firms (In Index): 508.

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.[34] Both of these indices are value-weighted, so moving from the 1000 to the 2000 significantly increases the share of passive ownership in a stock. The firm goes from being the smallest firm in an index of large firms, to the biggest firm in an index of small firms, increasing its relative weight by a factor of 10 (see e.g., Appel et al. (2016)).

The increase in passive ownership associated with S&P 500 index addition is not a perfect natural experiment because firms are not added at random. Once added, firms receive

---

[34]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.

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[35]. In these papers, the change in institutional 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. Using my Russell experiment to create an alternative difference-in-differences design yields 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 11 compares the level of passive ownership around the index rebalancing date between the treated and control group. Passive ownership starts increasing in the treated

---

[35]See 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). Most of these papers are using Russell reconstitution to identify changes in institutional ownership, rather than passive ownership.

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.
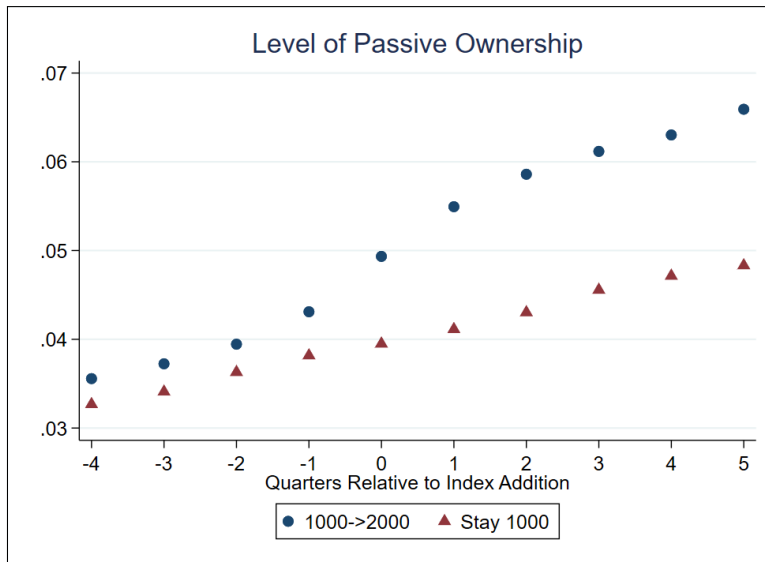

Figure 11. Russell 1000/2000 Reconstitution and Changes in Passive Ownership. Average level and increase in passive ownership for control firms and firms moved from the Russell 1000 to the Russell 2000. Control firms are all firms in the Russell 3000 ranked 900 to 1100 that did not move from the 1000 to the 2000 or from the 2000 to the 1000.

For the Russell experiment, I use the same difference-in-differences 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 8 contains the regression results. For comparison, I included a row with the reduced form estimates, which correspond to the value-weighted specification with all controls and fixed effects estimated in Section 4. Like the S&P 500 results, the estimated coefficients from the heterogeneous treatment intensity specification are similar in magnitude to the cross-sectional estimates for the volume and drift columns.

The results for earnings day volatility have the same sign, but are insignificant. The primary reason for this is 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. There could be many explanations for this, including non-fundamental volatility from increased ETF ownership (see e.g., Ben-David et al. (2018)) or the negative correlation between returns and volatility. The weakness of the volatility results could also be due to the relatively short sample period (1996-2012), and the smaller number of treated and control firms than the S&P 500 experiment.

	Volume	Drift	Volatility
Binary Treatment			
Treated	-1.024**	-0.00622**	0.000223
	(0.483)	(0.003)	(0.009)
R-squared	0.127	0.158	0.073
Continuous Treatment			
Interaction	-22.57**	-0.114*	0.0199
	(9.178)	(0.060)	(0.197)
R-squared	0.129	0.155	0.073
Reduced Form	-23.96***	-0.0965***	0.381**

Table 8 Effects of Russell 1000/2000 Index Reconstitution. Estimates from:

$$\Delta Outcome_{i,t} = \alpha + \beta \times Treated_{i,t} + FE + \epsilon_{i,t}$$

And:

$$\Delta Outcome_{i,t} = \alpha + \beta \left(\Delta Passive_{i,t} \times Treated_{i,t} \right) + FE + \epsilon_{i,t}$$

Where $Treated_{i,t}$ is a dummy variable equal to one if a firm switched from the Russell 1000 to the Russell 2000. Both specifications include month of index addition fixed effects. There are 216 treated firms and 158 control firms. Standard errors in parenthesis.

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 the Online Appendix, I show that this treatment effect is washed out by the time trend toward increased passive

ownership.[36].

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 re-run the cross-sectional 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 15% 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. 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.

---

[36]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.

A   Appendix A: Model details

A.1   Prices, Demands and Posteriors

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

Define Q as: $\frac{1}{\rho} \times \phi \times (S)^{-1}$, where ϕ is the share of rational traders who decide to become informed at cost c. The price function is:

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

$$A_3 = \frac{1}{\rho}\left((V)^{-1} + Q * (U)^{-1} * Q + Q\right)$$

$$A_0 = \frac{1}{\rho}A_3^{-1}\left((V)^{-1}\mu + Q(U)^{-1}\overline{\mathbf{x}}\right) \tag{30}$$

$$A_1 = A_3^{-1}\left(Q + \frac{1}{\rho}Q(U)^{-1}Q\right)$$

$$A_2 = A_3^{-1}\left(\mathbf{I}_n + \frac{1}{\rho}Q(U)^{-1}\right)$$

The demand functions for informed/uninformed investors are:

$$\text{Uninformed: Demand} = G_0 + G_{2,un}\mathbf{p}$$

$$\text{Informed, investor } j\text{: Demand} = G_0 + G_1\mathbf{s_j} + G_{2,inf}\mathbf{p} \tag{31}$$

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

$$\gamma = \rho\left(A_2^{-1} - Q\right)$$

$$G_0 = A_2^{-1}A_0$$

$$G_{2,un} = \frac{1}{\rho}\gamma$$

$$G_{2,in} = \frac{1}{\rho}\left(\gamma + S^{-1}\right) \tag{32}$$

$$G_1 = \frac{1}{\rho}S^{-1}$$

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:

$$
\begin{aligned}
E_{1,j}[\mathbf{z}|\mathbf{s_j}, \mathbf{p}] &= B_{0,in} + B_{1,in}\mathbf{s_j} + B_{2,in}\mathbf{p} \\
V_{in}^a &= \left(V^{-1} + QU^{-1}Q + S^{-1}\right)^{-1} \\
B_{0,in} &= \rho V_{in}^a G_0 \\
B_{1,in} &= \rho V_{in}^a G_0 \\
B_{2,in} &= \mathbf{I}_n - \rho V_{in}^a G_{2,in}'
\end{aligned}
\tag{33}
$$

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

$$
\begin{aligned}
E_{1,j}[\mathbf{z}|\mathbf{p}] &= B_{0,in} + B_{2,un}\mathbf{p} \\
V_{un}^a &= \left(V^{-1} + QU^{-1}Q\right)^{-1} \\
B_{0,un} &= \rho V_{un}^a G_0 \\
B_{2,un} &= \mathbf{I}_n - \rho V_{un}^a G_{2,un}'
\end{aligned}
\tag{34}
$$

A.2 Baseline parameters

Table 9 contains the baseline parameters. I take most of them from Kacperczyk et al. (2016). For a discussion of the differences between my parameter choices and theirs, see the Online Appendix.

Mean asset payoff	a_i	15
Volatility of idiosyncratic shocks	σ_i^2	0.55
Volatility of noise shocks	σ_x^2	0.5
Risk-free rate	r	1
Initial wealth	w_0	220
Baseline Learning	α	0.001
# idiosyncratic assets	n	8
Total supply of idiosyncratic assets	$\overline{x}$	20

Table 9 Baseline Parameters.

A.3 Numerical method for solving the model

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)[37].

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.

B Appendix B: Empirical Details

B.1 Decomposition of earnings-days' share of annual volatility

Figure 12 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.

B.2 Levels vs. First Differences

All the cross-sectional results, and the quasi-experimental results in Section 5, are robust to running the regressions in levels rather than first differences. A comparison between the levels and first-differences results are in Table 10. The levels results are stronger for the drift and volatility measures, both with larger point estimates and becoming more statistically significant. The results are weaker for volume, although they are still economically large and statistically significant at the 1% level.

---

[37]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.
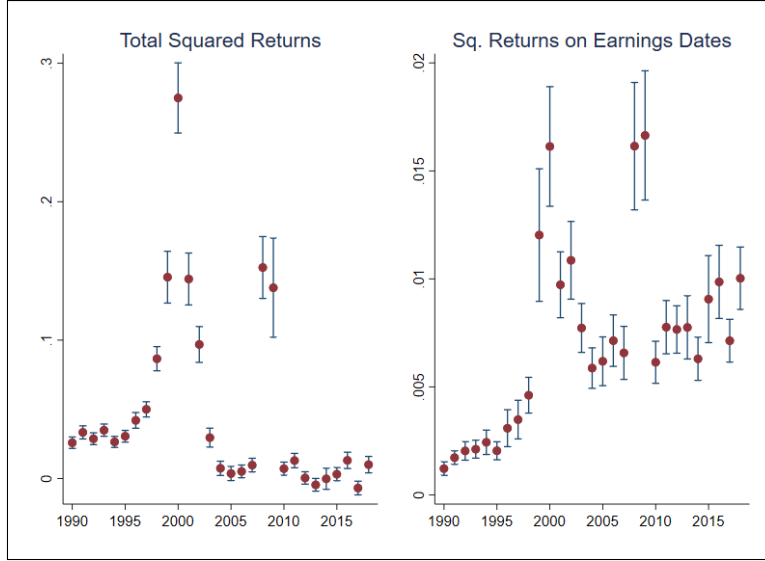
Figure 12. Decomposition of QVS. This figure plots coefficients from a regression of the numerator and denominator of QVS on a set of year dummy variables. For firm i in year t the quadratic variation share (QVS) is defined as: $QVS_{i,t} = \sum_{\tau=1}^{4} r_{i,\tau}^2 / \sum_{j=1}^{252} r_{i,j}^2$, 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.

The levels and first differences regressions have different interpretations. For example: In Equation 23, with firm and time fixed effects, $\beta < 0$ has the following interpretation: *Relative to firm-level average changes in price informativeness*, within in a particular year/quarter, firms which have larger increases in passive ownership have larger decreases in price informativeness. In levels, $\beta < 0$ would have a slightly different interpretation: *Relative to firm-level average price informativeness*, within in a particular year/quarter, firms which have higher passive ownership have lower price informativeness.

Running the regression in first differences seems inconsistent with the model, which is about the *level* of passive ownership. Using differences, however, will prevent bias arising from the persistence of passive ownership at the firm-level. Empirically, passive ownership at the firm level follows a process like: $Passive_{i,t} = Passive_{i,t-1} + \gamma_{i,t} AggregatePassive_t + e_{i,t}$, where $\gamma_{i,t}$ is how sensitive firm i is to changes in aggregate passive ownership at time t. If a

	Volume	Drift	Volatility
Levels	-7.158***	-0.0457***	0.116**
	(2.001)	(0.011)	(0.052)
First Diffs.	-15.99***	-0.0323**	0.106*
	(3.101)	(0.014)	(0.056)
Firm/Time FE	Yes	Yes	Yes
Controls	Yes	Yes	Yes
Weight	Eq.	Eq.	Eq.

Table 10 Cross-Sectional Results: Levels vs. First Differences. The first row contains the estimates from the regression in levels: $Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$ while the second row has estimates from the regression in first-differences:
$\Delta Outcome_{i,t} = \alpha + \beta \Delta Passive_{i,t} + \gamma X_{i,t-1} + \text{Fixed Effects} + e_{i,t}$
Both contain time and firm fixed effects. The levels regression has standard errors double clustered at the firm/time level. The first differences regression has standard errors computed using panel Newey-West, with lags equal to 1.5x the number of overlapping observations.

firm had a large increase in passive ownership between 1990 and 2010, it will likely have a higher than average level of passive ownership for the rest of the sample (2011-2018), even if $\gamma_{i,t}$ decreases. The bias in a levels regression would be even larger if there is persistence in price informativeness at the firm level.

The Online Appendix simulates the distribution of β for levels and first differences under several different scenarios. The distributions of β in levels and differences are different when the true $\beta = 0$: The spread of β's is wider with levels, which could lead to more false positives in terms a relationship between the two quantities. If there is a strong relationship between passive ownership and price informativeness, the distributions of β are almost identical in levels and first-differences.

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.

Appel, I. R., Gormley, T. A., and Keim, D. B. (2016). Passive investors, not passive owners. *Journal of Financial Economics*, 121(1):111–141.

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.

Bertrand, M., Duflo, E., and Mullainathan, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly journal of economics*, 119(1):249–275.

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.

Breugem, M. and Buss, A. (2019). Institutional investors and information acquisition: Implications for asset prices and informational efficiency. *The Review of Financial Studies*, 32(6):2260–2301.

Buss, A. and Sundaresan, S. (2020). More risk, more information: How passive ownership can improve informational efficiency.

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.

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.

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*.

Khan, M., Srinivasan, S., and Tan, L. (2017). Institutional ownership and corporate tax avoidance: New evidence. *The Accounting Review*, 92(2):101–122.

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)*.

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.

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.