Passive Ownership and Price Informativeness

MARCO SAMMON *

August, 2021

ABSTRACT

I propose three new empirical measures of price informativeness motivated by a theoretical model. I find average price informativeness declined over the past 30 years and passive ownership is 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. The model offers two explanations for the empirical results: passive ownership decreases investors' incentives to become informed and leads investors to re-allocate their attention from stock-specific risks to systematic risk.

Keywords: Passive ownership, Price informativeness. JEL classification: G12, G14.

*Harvard Business School. Online appendix and replication code. I would like to thank my dissertation committee: Scott Baker, Alex Chinco, Ravi Jagannathan, Robert Korajczyk, and Dimitris Papanikolaou. I would also like to thank Nick Bloom, Ian Dew-Becker, Simon Gloßner, Robin Greenwood, Sam Hanson, Martin Thyrsgaard, Luis Viceira and participants at the Kellogg Bag Lunch for helpful comments. The Avi Nash Fund provided financial support.

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 the incorporation of information into stock prices?

There is no consensus answer to this question in the theoretical or applied literature. For example: Cong et al. (2020) and Glosten et al. (2021) argue that passive ownership can increase the incorporation of systematic news for stocks with otherwise weak information environments. On the other hand, Ben-David et al. (2018) and Kacperczyk et al. (2018) provide evidence that passive ownership increases non-fundamental volatility and can lead investors to trade less aggressively on their private information.

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 contain less information. I use the term *price informativeness* to broadly capture the notion of how well a stock's price reflects information about the fundamental value of the firm. I document a new stylized fact consistent with this standard 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 turnover (right panel) around earnings announcements. The sample includes firms in the top decile of standardized unexpected earnings (SUE) each quarter 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 slowdown 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 earnings day return 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 contained a larger share of the earnings information in the early 1990s, when passive

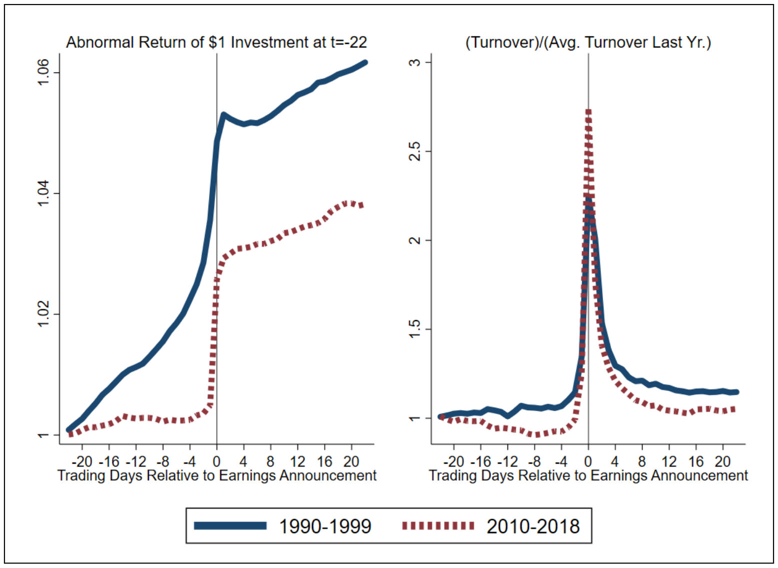

Figure 1. Returns and turnover around earnings announcements. Each plot represents the cross-sectional average for firms in the top decile of Standardized Unexpected Earnings (SUE), defined as: $SUE_{i,t} = \frac{Earnings_{i,t} - Earnings_{i,t-4}}{\sigma_{(t-1,t-8)}(Earnings_{i,t} - Earnings_{i,t-4})}$. Deciles of SUE are calculated each quarter. Abnormal returns are returns minus the returns on the CRSP value-weighted index. Turnover is total trading volume divided by shares outstanding. Abnormal turnover is turnover divided by the firm-level average turnover over the past 252 trading days.

ownership was negligible, than they do now, when passive ownership is large.

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.

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.*[1]" 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

[1]Daniel Gamba, Global Head of Active Equity Product Strategy, BlackRock, quoted in Balchunas (2016).

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.*[2]" 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. An empirical analogue is a regression of future fundamentals on current prices:

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

Larger values of $\hat{\beta}$ (scaled by the standard deviation of prices) 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. 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 system-

[2]Bruno del Ama, former CEO of Global X, quoted in Balchunas (2016).

atic 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: trading volume, returns and volatility. I use these predictions to define three measures of price informativeness: (1) pre-earnings volume, (2) pre-earnings drift and (3) earnings-day volatility.

Through simulations, I show why these quantities are useful measures of price informativeness: As the share of informed investors decreases, or attention on stock-specific risks decreases, pre-earnings trading volume declines, the pre-earnings drift declines and earnings-day volatility increases. 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 passive ownership is associated with lower turnover before earnings announcements. Passive ownership is also correlated with lower pre-earnings drift and higher 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., Regulation Fair Disclosure, passed in August 2000 and changes in the enforcement of insider trading laws, as discussed in Coffee (2007)), all the cross-sectional regressions include year/quarter fixed-effects. 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. This relationship, however, is not necessarily casual. In fact, there are two forces that can increase passive ownership in the model: (1) A 'supply side' effect, where a technological improvement leads to an increase in passive ownership, which changes investors' learning behavior and (2) A 'demand side' effect, where a change in investors' preferences (e.g., an increase in risk aversion) leads to a change in learning, which increases demand for the ETF and increases equilibrium passive ownership. Outside the model, it's possible that 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 increases in passive ownership that are plausibly uncorrelated with firm fundamentals. To this end, I design two natural experiments based on 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.

1.1 Related literature

My paper contributes to several strands of literature. The first is the ongoing debate on the effect of passive ownership on price informativeness. Several papers argue that rising passive ownership and the introduction of ETFs can improve price informativeness. Buss and Sundaresan (2020) show that passive ownership makes prices more informative through its effect on firms' investment behavior and the resulting endogenous change in investors' learning choices. In Ernst (2020), ETFs make prices more informative, as they offer an additional way to trade information into stock prices. Lee (2020) shows how passive investors can increase the benefit of being informed by increasing liquidity. Malikov (2020) shows the conditions under which falling information costs can lead to simultaneous increases in passive ownership and price informativeness. Beschwitz et al. (2020) argue that ETFs make it easier to short constituent stocks, also making prices more informative. Finally, Glosten et al.

(2021) provide theoretical and empirical evidence that passive ownership makes prices more informative by increasing the incorporation of systematic information. While introducing and growing the ETF in my model can make individual stock prices more informative, it is through a new mechanism: Growing passive ownership allows investors to take more aggressive bets on stock-specific information, because they can use the ETF to hedge their exposure to systematic risk.

Other papers have argued that passive ownership decreases price informativeness. Garleanu and Pedersen (2018) preset a model where declining costs of passive management make prices less informative by decreasing incentives to gather information. Kacperczyk et al. (2018) show that passive ownership can decrease how aggressively informed investors trade on their private information. Breugem and Buss (2019) develop a model where benchmarking makes prices less informative by reducing the returns to private information. Haddad et al. (2021) show through a model that increasing passive ownership lowers the aggregate responsiveness of investor demand to prices, which makes them less informative. My model builds on these results by showing that (1) passive ownership can have different effects on investors' incentives to gather information on systematic risk vs. idiosyncratic risk and (2) which of these effects dominate depends on investors' risk tolerance.

Finally, in some papers, the effect is ambiguous or neutral. Cong et al. (2020) present a model where introducing an ETF can increase the incorporation of factor-specific information, especially in otherwise illiquid stocks, while stock-specific information declines. Bhattacharya and O'Hara (2018) also present a model where ETFs can lead to increased incorporation of systematic information, at the expensive of stock-specific information. Coles et al. (2020) present an irrelevance result where passive ownership has no effect on price informativeness, and present empirical evidence consistent with this prediction. I add to this literature by illustrating alternative competing forces that can lead passive ownership to have an ambiguous effect not only on the overall incorporation of information into prices, but even the stock-specific component of information.

My paper also contributes to the literature on the effect of composite securities (e.g., ETFs) on asset markets. In both Subrahmanyam (1991) and Gorton and Pennacchi (1993), introducing a futures contract benefits uninformed investors by reducing the informed traders' information advantage. In my model, the ETF makes uninformed investors better off because initially they cannot perfectly diversify away stock-specific risk. Subrahmanyam (1991) shows that introducing a futures contract can increase the incorporation systematic information, and decrease the incorporation of stock-specific information via endogenous changes in learning. While I have similar learning channels, they are driven by standard portfolio considerations, rather than the behavior of liquidity traders: Risk aversion determines whether investors use the ETF to hedge systematic risk when betting on individual stocks or execute a market timing strategy.

More recently, Ben-David et al. (2018) show that ETFs can increase non-fundamental volatility by transmitting shocks in the ETF to the underlying securities. Chinco and Fos (2021) provide evidence that ETF rebalancing cascades can generate noise in financial markets. I add to this literature through evidence on a new channel: ETFs can affect trading volume and volatility through their effect on investors' learning.

Finally, I contribute to the literature on measuring price informativeness and trends in price informativeness. Both Bai et al. (2016) and Dávila and Parlatore (2018) document increases in average price informativeness over the past 30 years. Farboodi et al. (2020) show that these increases in price informativeness are mainly concentrated in large/growth stocks. I offer an alternative strategy for measuring price informativeness that exploits the timing of when information is released, and does not rely on knowing how to measure the fundamental value of a firm.

By focusing on earnings announcements, and using the readily observable quantities that come out of the model, I come to the opposite conclusion: average price informativeness has decreased over the past 30 years. Although these results are contradictory, I am making a narrower claim: my results are focused on how well prices reflect information contained in earnings announcements just before that earnings information is released to the public.

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) the modeling of systematic risk.

Investors

There are a unit mass of rational investors which fall into two groups: informed and uninformed. At time 1, they both have CARA preferences over time 2 wealth. Informed investors receive signals at time 1 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:

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

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.[3] Each stock has $\overline{x}_i$, shares outstanding and noise trader demand shocks

[3]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. In this case, the number of risks would be equal to the number of assets. Without a systematic risk-factor, however, there is no guarantee the learning technology will be comparable between economies when the ETF i.e., the $n+1^{th}$ asset is and is not present. The Online Appendix subsection titled "Equivalence of Learning Technologies Between Rotated and Unrotated Versions of the Model" discusses this representation issue in detail.

$x_i \overset{\text{iid}}{\sim} N(0, \sigma^2_{i,x})$. 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, $\sigma^2_{i,x} = \sigma^2_x$. This assumption is not needed, but it simplifies the intuition for the key learning trade-offs. For an extension where individual stocks load differently on systematic risk, and have heterogeneous volatility of their idiosyncratic risk-factors, 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 parameterization I set $n = 8$.[4] Without this assumption, introducing the ETF would have no effect on investors' behavior: The lack transaction costs means that as n goes to infinity, the individual investors could perfectly replicate the systematic risk factor on their own without the ETF.

Signals

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

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

where $\epsilon_{i,j} \overset{\text{iid}}{\sim} N(0, \sigma^2_{\epsilon_{i,j}})$, $\epsilon_{f,j} \sim N(0, \sigma^2_{\epsilon_{f,j}})$ 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:

$$\sigma^2_{\epsilon_{i,j}} = \frac{1}{\alpha + K_{i,j}}, \quad \sigma^2_{\epsilon_{f,j}} = \frac{1}{\alpha + K_{n+1,j}} \tag{3}$$

[4]With $n = 8$, consider the case where $\sigma_f = 0.25$ and $\sigma = 0.55$: Then, 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.

where $\alpha > 0$.[5] *Baseline learning*, 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 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 parameterization, 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, rather than pay for a fixed K.

Portfolio Choice

Define terminal wealth:

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

where $w_{0,j}$ is initial wealth, c is the cost of becoming informed, $\mathbf{z}$ 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:

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

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

[5] 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. 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 assets.

because investors do not independently receive information about the $n+1^{th}$ risk-factor. Because there are more risks than independent signals/stocks, investors cannot rotate the economy to think in terms of synthetic assets exposed only to risk-factor payoffs, rather than stock payoffs (see e.g., Veldkamp (2011)). 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).[6]

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 $S_j = S$ and $K_{i,j} = K_i$ for all j. Given the learning technology, S^{-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 equations for equilibrium prices, beliefs and demands come directly from Admati (1985) (see Online Appendix for details):

$$\mathbf{p} = A_0 + A_1\mathbf{z} - A_2(\overline{\mathbf{x}} + \mathbf{x}) \tag{6}$$

Beliefs

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

$$s_p = A_1^{-1}\left(p - A_0 + A_2(\overline{\mathbf{x}} + \mathbf{x})\right) \tag{7}$$

[6]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.

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

Demands

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

$$\begin{aligned} \text{Uninformed: Demand}{=}G_0 + G_{2,un}\mathbf{p} \\ \text{Informed, investor } j\text{: Demand}{=}G_0 + G_1\mathbf{s_j} + G_{2,inf}\mathbf{p} \end{aligned} \tag{8}$$

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

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:

$$U_{0,j} = E_0\left[E_{1,j}[w_{2,j}] - 0.5\rho Var_{1,j}[w_{2,j}]\right] \tag{9}$$

because time two wealth is normally distributed. The Online Appendix discusses how Equation 9 is derived from Epstein and Zin (1989)-style preferences, and how this differs 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. With the ETF present, the model is similar to Kacperczyk et al. (2016), with the key difference being an endogenously supplied ETF.

Asset payoffs

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

$$z_{n+1} = a_{n+1} + f \tag{10}$$

When it is present, the ETF initially has average supply $\overline{x} = 0$, 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 (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.

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. The Online Appendix contains a more thorough discussion of the implications of these two assumptions on the equilibrium size of the ETF.

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

$$U_{0,j} = E_0 \left[E_{1,j}[w_{2,int}] - 0.5\rho^{int} Var_{1,j}[w_{2,int}] \right] \tag{11}$$

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 owned by the ETF. This maps almost exactly to the definition of passive ownership in the empirical exercises, which is the fraction of each stock's shares outstanding owned by all passive funds.

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

$$w_{2,int} = v \left(\underbrace{\sum_{i=1}^{n} (z_i - p_i)}_{\text{Idio. Risk}} - \underbrace{n(z_{n+1} - p_{n+1})}_{\text{Sys. Risk}} \right) \tag{12}$$

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, this assumption is designed 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 to 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' demands, 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:

$$\begin{aligned} s_{i,j} = a_i + (f + \epsilon_{f,j}) + (\eta_i + \epsilon_{i,j}) \text{ for } i = 1, \ldots, n \\ s_{n+1,j} = a_i + (f + \epsilon_{f,j}) \end{aligned} \tag{13}$$

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 expected 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. (2021), 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...*[7]*"*

[7] Quoted in Balchunas (2016).

Another feature it captures is that ETFs have made it easier to hedge out/short systematic risk. According to Goldman Sachs Hedge Fund Monitor (2016), *"ETFs account for 27% of hedge funds' short equity positions."* 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. Finally, ETFs cover more sectors/indexes than futures contracts and mutual funds.

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. The Online Appendix explains how I use these two conditions to numerically solve the model.

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 the systematic risk-factor (σ_f^2) is low, relative to the idiosyncratic risk-factors (σ^2). This difference in volatilities means that there are more profit opportunities in the stock-specific risk factors 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. This is because they cannot perfectly hedge the exposure to systematic risk embedded in any given stock. The Online Appendix presents two asset examples that illustrate 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. Each example shares the baseline parameters in Table 1, which are mostly borrowed from Kacperczyk et al.

(2016). These examples/parameters are not a calibration and are only designed to illustrate the model's learning mechanisms.

Model Object	Symbol	Value
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
# of idiosyncratic assets	n	8
Total supply of idiosyncratic assets	$\overline{x}$	20

Table 1 Baseline parameters. Parameters shared across the intensive/extensive learning margin examples.

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 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.

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

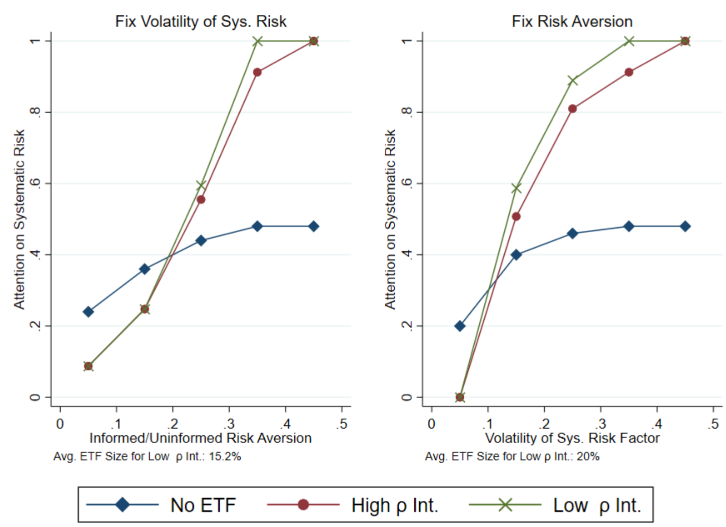

Figure 2. Effect of the ETF on intensive learning margin. In both panels, the share of informed investors is fixed at 60%. In the left panel, $\sigma_f = 0.35$. In the right panel, $\rho = 0.35$. The y-axis reports the share of investors' attention devoted to the systematic risk-factor.

larger the ETF is, the cheaper it is to hedge systematic risk. The size of the ETF affects the cost of hedging systematic risk because the ETF trades at a premium to an equal-weighted basket of all the underlying stocks. Investors are risk averse, so they are willing to pay a premium for a security which is not exposed to idiosyncratic risk. As the ETF becomes larger, this ETF premium shrinks.

The hedging channel will show up in investors' demand functions. For informed investors, G_1 from Equation 8 is a measure of how sensitive demand is to their private signals. Table 2 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.

Because all the stocks have the same expected 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.

		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 2 Effect of the ETF on informed investors' demand functions. 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.

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 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.

Risk-bearing capacity is designed to capture the following intuition: In the model, 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.

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.

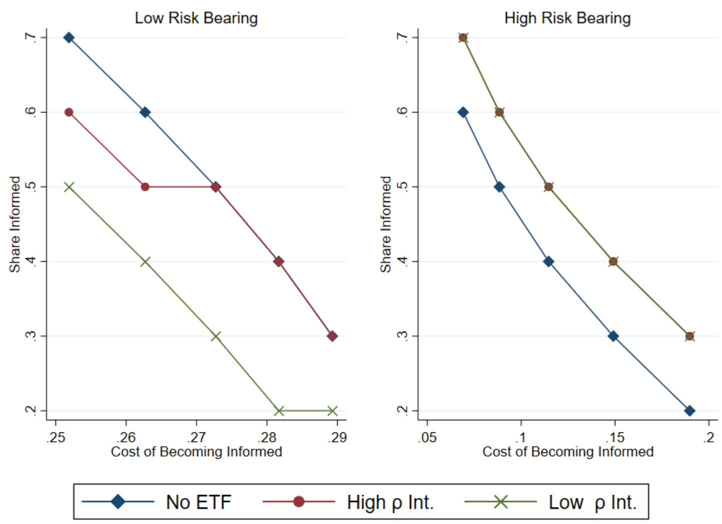

Figure 3. Effect of the ETF on extensive learning margin. Left panel: $\sigma_f = 0.25$, $\rho = 0.15$. Right panel: $\sigma_f = 0.05$, $\rho = 0.05$. The x-axis is the cost in dollars of becoming informed. The y-axis reports the share of investors who become informed in equilibrium at this cost.

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 Defining price informativeness

The natural first step in relating passive ownership to price informativeness 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. Motivated by an alternative model, Dávila and Parlatore (2018) and Dávila and Parlatore (2021) measure price informativeness as the variance of prices, conditional on fundamentals. These conflicting results suggest that while it may be straightforward to measure price informativeness *within* a given model, it does not mean that measure will be valid *across* models.

The correct measure of future fundamentals is also not obvious: In a static model like Grossman-Stiglitz, there are a finite number of cash flows, but in reality, firms are long-lived. Maybe fundamentals should be defined as *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)).

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

Pre-Earnings Trading Volume

Although the model features a continuum of investors, when simulating the economy, there are a finite number, which I set to 10,000. At $t = 0$, I assume all of the investors are endowed with $1/10,000^{th}$ of $\overline{x}$. 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, adjusting for investor j's share of the noise shock and final holdings.

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

$$\sum_{j}^{J} \left| \mathbf{q}_j - \left(\overline{\mathbf{x}} + \mathbf{x}\right)/(J) \right| \tag{14}$$

where the first term $\mathbf{q}_j$ is investor j's demand, and the second term $\left(\overline{\mathbf{x}}+\mathbf{x}\right)/(J)$ 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:

$$DM = \begin{cases} \frac{1+r_{(0,1)}}{1+r_{(0,2)}} & \text{if } r_2 > 0 \\ \frac{1+r_{(0,2)}}{1+r_{(0,1)}} & \text{if } r_2 < 0 \end{cases} \tag{15}$$

where $r_{(0,t)}$ is the cumulative market-adjusted return from 0 to t.[8] 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. 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:

$$\frac{r_2^2}{r_1^2 + r_2^2} \tag{16}$$

[8] I work with market-adjusted returns to account for the effect of passive ownership on the market risk premium. 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 the risk premium.

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

2.5 Effects 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. I show that changes in the intensive and extensive learning margins have unambiguous effects on all three price informativeness measures.

2.5.1 Effect of Learning on Price Informativeness: Directly varying attention and the share of informed investors

To understand the relationship between the intensive learning margin and the price informativeness measures, Panel A of Figure 4 fixes the share of informed investors at 50%, and directly varies attention on systematic risk. As expected, increasing total attention on stock-specific risks increases pre-earnings trading volume, increases the pre-earnings drift and decreases the share of volatility on earnings days. All of these plots are consistent with increased attention on stock-specific risks leading to increased price informativeness.

To understand the relationship between the extensive learning margin and price informativeness, Panel B of Figure 4 fixes investors attention on systematic risk at 50% and directly varies the share of informed investors. Increasing the share of informed investors increases pre-earnings trading volume, increases the pre-earnings drift and decreases volatility on earnings days. These results are consistent with an increased share of informed investors leading to increased price informativeness.

2.5.2 Effect of Learning on Price Informativeness: Indirectly varying attention

When directly varying the share of informed investors and/or investors' attention on systematic risk, there is no guarantee that these learning choices are optimal. To ensure

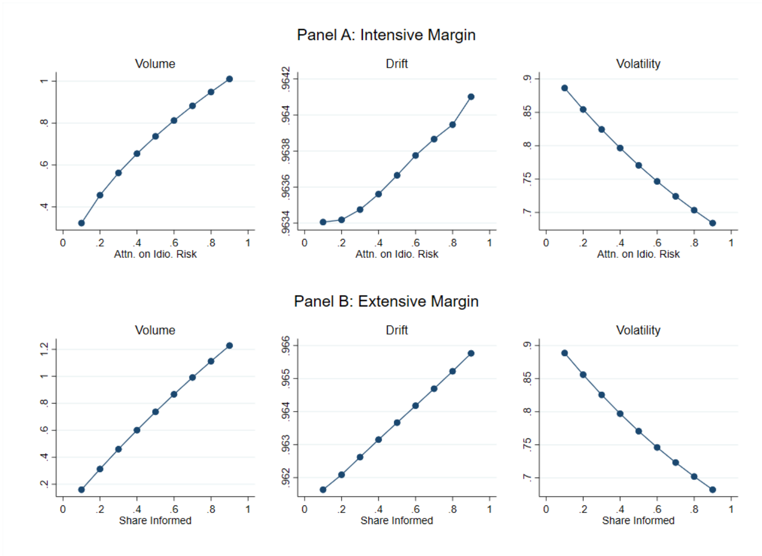

Figure 4. Effect of learning on price informativeness measures. Panel A plots the three measures of price informativeness against attention on systematic risk, fixing the share of informed investors at 50%. Panel B plots the three measures of price informativeness against the share of informed investors, fixing attention on systematic risk at 0.5. In all panels, $\rho = 0.15$, $\sigma_f = 0.3$, and the risk aversion of the ETF intermediary, ρ^{int}, is set to infinity so they create no shares of the ETF.

that the relationship between investors' leaning behavior and price informativeness also holds at optimal learning choices, 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, while ensuring that attention is still optimally allocated.

Panel A of Figure 5 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.

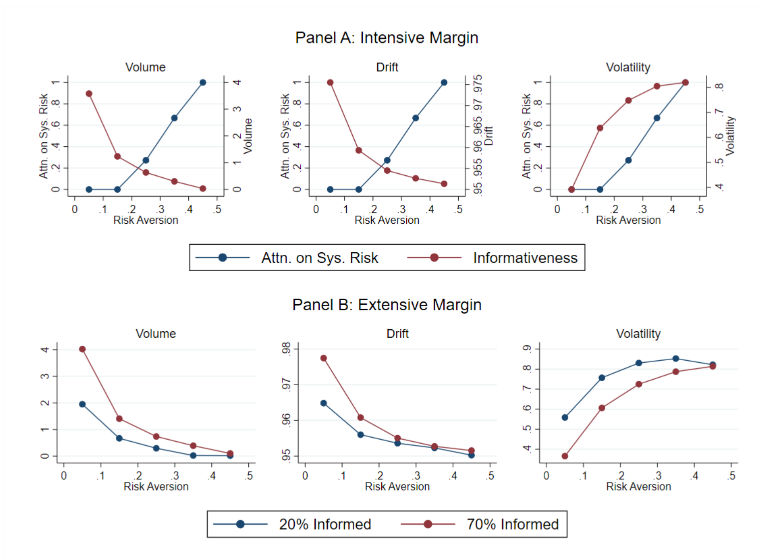

Figure 5. Effect of learning on price informativeness measures. Panel A plots the attention on systematic risk and three measures of price informativeness against informed and uninformed investors' risk aversion, ρ. Panel B plots the three measures of price informativeness against ρ for high and low 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, ρ^{int}=1. In Panel A the share of informed investors is set to 60%.

Figure 5 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.

2.6 Effect of passive ownership on price informativeness

Between 1990 and 2018, passive ownership grew from nothing to owning 15% of the US stock market. In Figure 6, 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. All the price informativeness measures are only calculated for the stocks i.e., assets 1 to n.

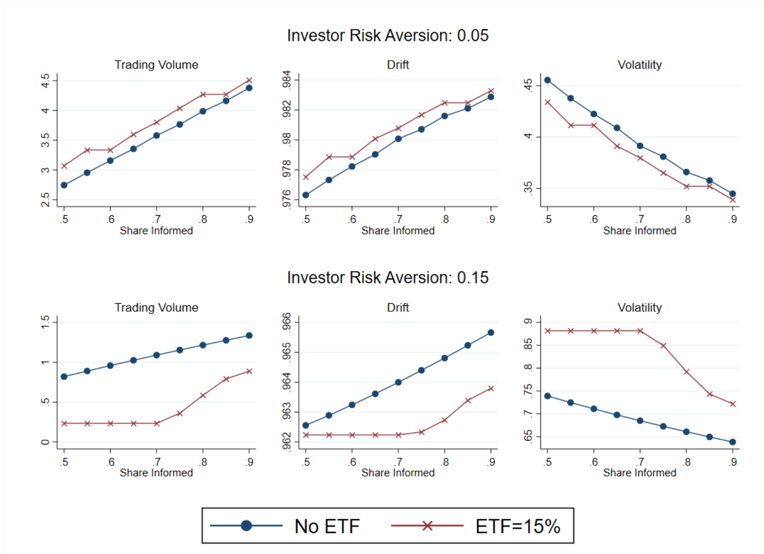

Figure 6. Effect of passive ownership on price informativeness measures. Top panels: $\rho = 0.05$, $\sigma_f = 0.15$. Bottom panels: $\rho = 0.15$, $\sigma_f = 0.35$. The cost of being informed is set so to match the share of investors who become informed when the ETF is not present.

The top 3 panels are averages of the price informativeness measures in an economy with low risk aversion, $\rho = 0.05$, and low volatility of the systematic risk factor $\sigma_f = 0.15$. Increasing the share of informed investors (moving to the right along the x-axis) unambiguously increases price informativeness: it increases the pre-earnings drift, decreases earnings-day volatility and increases pre-earnings trading volume. In this economy, growing passive ownership increases price informativeness: it increases pre-earnings trading volume, increases the pre-earnings drift and decreases the volatility on earnings announcement dates. 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 higher risk aversion $\rho = 0.15$ and higher volatility of the systematic risk factor $\sigma_f = 0.35$. Volume, drift and volatility all suggest that growing passive ownership leads to less informative prices. This is evidence of an economy where the hedging channel is dominated by the diversification and market timing channels.

Figure 6 shows that as with the extensive and intensive learning margins, the net 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 funds with names that identify them as index funds, according to the criteria in Appel et al. (2016). Index funds are identified using the index fund flag in the CRSP mutual fund data. 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.

There is, however, an important distinction between the variation in passive ownership in the data (cross-sectional) and in the model (across economies). To model this, there would need to be stocks not held by the ETF. While this is possible, it would complicate the solution method. Specifically, it would add additional dimensions to the attention allocation problem e.g., between the idiosyncratic risk-factors of stocks inside and outside the index.

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 7 shows that passive ownership increased from almost zero in 1990, to now owning about 15% of the US stock market. As of 2018, passive ownership was over 40% of total mutual fund and ETF assets.

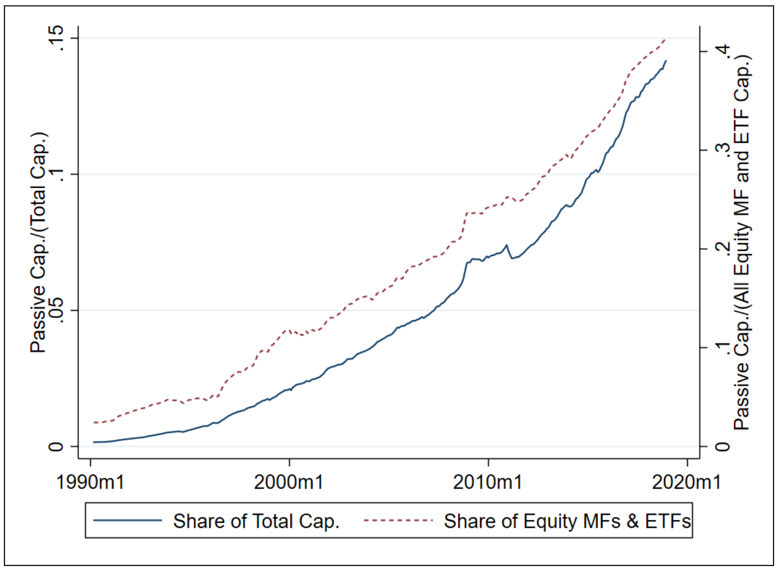

Figure 7. The rise of passive ownership: 1990-2018. Passive funds are defined as all index funds, all ETFs, and all mutual funds with names that identify them as index funds (see e.g., Appel et al. (2016)). Total equity mutual fund and ETF assets is the sum of all stock holdings in the Thompson S12 data that can be matched to CRSP.

This definition likely understates the true level 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 (2019), 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 Friday, 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.[9] 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 not only by the model, but also by the literature on asymmetric information (see e.g., Akerlof (1970), Milgrom and Stokey (1982)). As information asymmetries become larger, uninformed investors are less willing to trade because of adverse selection: They are concerned that the only people willing to trade with them are better informed, so any trades they make are guaranteed to be bad deals. In the stock 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. For example: if all investors are uninformed, there would be no trade, as they would all have the same posterior beliefs about every security. This is why increasing the share of informed investors or increasing investors' attention on stock-specific risks unambiguously increases trading volume before the earnings announcement date i.e., at $t = 1$.

Unlike in the model, there are many dates between earnings announcements. The model's predicted change in trading volume may be spread out over the month (22 trading days)

[9]All results are similar when using Compustat analogues: (1) Report Date of Quarterly Earnings (RDQ) instead of the IBES announcement date, and (2) Diluted Earnings Per Share Excluding Extraordinary Items (EPSFXQ) instead of IBES earnings value.

before an earnings announcement. Let t denote an effective earnings announcement date. Define turnover T as total daily volume for stock i in CRSP divided by shares outstanding. Then, define abnormal turnover for firm i, from time $t-22$ to $t+22$ as:

$$AT_{i,t+\tau} = \frac{T_{i,t+\tau}}{\overline{T_{i,t-22}}} = \frac{T_{i,t+\tau}}{\sum\limits_{k=1}^{252} T_{i,t-22-k}/252} \tag{17}$$

Where abnormal turnover, $AT_{i,t+\tau}$, is turnover divided by the historical average turnover for that stock over the past year.[10] I use abnormal turnover to account for differences across stocks and within stocks across time. Historical average turnover, $\overline{T_{i,t-22}}$, is fixed at the beginning of the 22-day window before earnings are announced to avoid mechanically amplifying drops in trading.

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

$$AT_{i,t+\tau} = \alpha + \sum_{k=-21}^{22} \beta_k \mathbf{1}_{\{\tau=k\}} + e_{i,t+\tau} \tag{18}$$

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 in my sample and a ± 22 day window around each earnings announcement. Abnormal turnover is Winsorized at the 1% and 99% levels by year.

I run this regression for three sample periods: (1) 1990-1999 (2) 2000-2009 (3) 2010-2018. Figure 8 plots the estimates of β_k for $k=-21$ to $k=-2$. The estimate for $k=-1$ is omitted as the coefficients are about $5\times$ as large as the coefficients for $k=-21$ to $k=-2$, which forces a change of scaling that makes the plot harder to interpret. For each day, the average abnormal turnover is statistically significantly lower in the third period, relative to the first period.

Figure 8 confirms that there has been a drop in trading volume throughout the month

[10] All results are robust to instead normalizing by the average turnover for that firm over the past quarter.

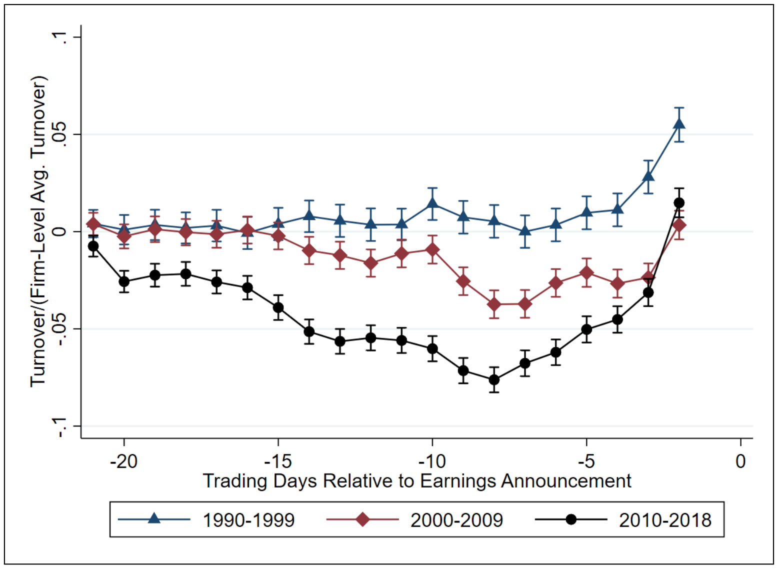

Figure 8. Decline of pre-earnings turnover. Plot of β_k estimated from the regression:

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

where $AT_{i,t+\tau}$, abnormal turnover, is turnover divided by the historical average turnover for that stock over the past year. $AT_{i,t+\tau}$ is Winsorized at the 1% and 99% level each year. Bars represent a 95% confidence interval around the point estimates. Standard errors are clustered at the firm level.

before each earnings announcement. Define cumulative abnormal pre-earnings turnover as:

$$CAT_{i,t} = \sum_{\tau=-22}^{-1} AT_{i,t+\tau} \tag{19}$$

the sum of abnormal turnover from $t-22$ to $t-1$ for firm i around earnings date t. $CAT_{i,t}$ is my first main empirical measures of price informativeness. Lower values of $CAT_{i,t}$ translate to less pre-earnings trading and in the model could be explained by a decrease in the share of informed investors and/or a decrease in attention to stock-specific risks. Between the 1990s and 2010s, average $CAT_{i,t}$ 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 (see e.g., Ball and Brown (1968), Foster et al. (1984), Weller (2018)). 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 excessive price impact, as in Kyle (1985).

Let $E_{i,t}$ denote earnings per share for firm i in quarter t in the IBES Unadjusted Detail File. Following Novy-Marx (2015), 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 9 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. The brown dashed line represents the average for firms with the most positive earnings surprises, while the blue dashed line represents the average for firms with the most negative earnings surprises. Between 2010 and 2018, firms in each decile move less before earnings days than between 2001 and 2007.

Figure 9'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. For an empirical analogue, define the pre-earnings drift for firm i as the cumulative market-adjusted gross return from $t-22$ to

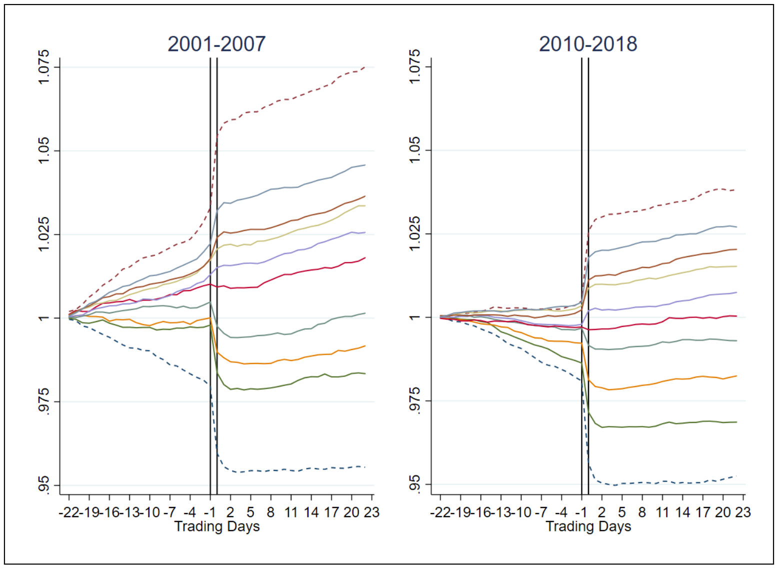

Figure 9. Decline of pre-earnings drift by SUE decile. Each quarter, I sort firms into deciles on standardized unexpected earnings. Each line represents the cross-sectional average market-adjusted return of \$1 invested at t=-22. The brown dashed line represents the average for firms with the most positive 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.

$t-1$, divided by the cumulative returns from $t-22$ to t, where t is an earnings announcement:

$$DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases} \tag{20}$$

The pre-earnings drift will be near one when the earnings day move is small relative to cumulative pre-earnings returns. $DM_{i,t}$ will be less than one when the earning-day return is large, relative to the returns over the previous 22 days. If r_t is negative, this relationship would be reversed, which is why the measure is inverted when r_t is less than zero. I work with gross returns, rather than net returns, to avoid dividing by numbers near zero.[11]

$DM_{i,t}$ is my second main empirical measure of pre-earnings price informativeness. Lower

[11]It has been well documented (see e.g., Mclean and Pontiff (2016), Martineau (2018)) 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 the numerator $1 + r_{(t-22,t)}$ with $1 + r_{(t-22,t+n)}$ for n between 1 and 5. All my results are qualitatively unchanged using these alternative pre-earnings drift measures which explicitly account for changes in post-earnings announcement returns.

values of $DM_{i,t}$ in the model can be explained by a lower share of informed investors and/or decreased attention on stock-specific risks. Average pre-earnings drift decreased by about -0.02 between 1990 and 2018.

3.5 Measure 3: Earnings days' share of 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 volatility occurring on earnings dates, relative to the volatility on the surrounding days.

Specifically, define the quadratic variation share (QVS) for firm i around earnings announcement τ as:

$$QVS_{i,\tau} = r_{i,\tau}^2 / \sum_{k=0}^{22} r_{i,\tau-k}^2 \tag{21}$$

where r denotes a market-adjusted daily return. The numerator is the squared earnings-day return, while the denominator is the sum of squared returns from $\tau - 22$ to τ. 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 announcement dates, we would expect larger values of QVS. In the model, an increase in QVS could be explained by a decrease in the share of informed investors and/or decreased attention to stock-specific risks. An advantage of QVS, relative to DM, is that it does not require knowing the sign of the earnings day return.

In this 23-day window (22 pre-earnings days + the earnings announcement itself), the earnings day is $1/23 \approx 4.3\%$ of observations, so values of $QVS_{i,t}$ larger than 0.043 imply that earnings days account for a disproportionately large share of total volatility.[12] Figure 10 plots

[12] All results are robust to defining QVS at the annual level i.e., defining the numerator to be 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*. They are also robust to including a post-earnings announcement window in the numerator i.e., defining $QVS_{i,\tau} = \left(\sum_{j=0}^{n} r_{i,\tau+j}^2 \right) / \left(\sum_{k=0}^{22} r_{i,\tau-k}^2 \right)$ for $n = 1, \ldots, 5$.

coefficients from a regression of QVS on a set of year dummy variables for all stocks in my sample. Average QVS increased from 7.7% in 1990 to 27.7% in 2018. The Online Appendix shows that the increase in QVS was due to a simultaneous increase in the numerator and a decrease in the denominator.

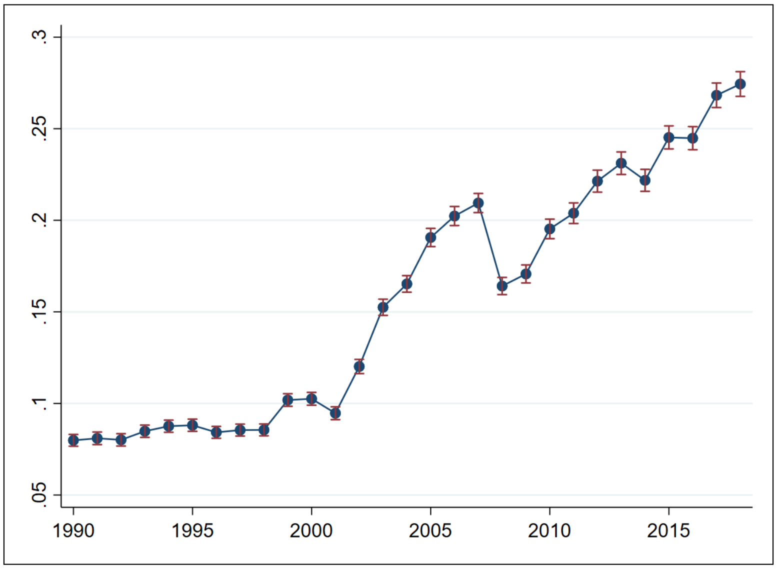

Figure 10. Increase in earnings-day volatility. This figure plots coefficients from a regression of QVS on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient. For firm i around earnings announcement τ the quadratic variation share (QVS) is defined as: $QVS_{i,\tau} = r_{i,\tau}^2 / \sum_{k=0}^{22} r_{i,\tau-k}^2$, where r denotes a market-adjusted daily return. The red bars 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: select the date 22 trading days before each earnings announcement – the start of each pre-earnings announcement window – to be a placebo earnings date. I then reconstruct the time-series averages of the pre-earnings turnover, 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 drift for the placebo earnings dates. Finally, there is no upward trend in the share of volatility on the placebo earnings dates. I repeat this exercise using randomly selected dates as placebo earnings announcements, and find no trends in any of the price informativeness measures. 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 scheduled Federal Open Market Committee (FOMC) meeting dates. As shown in Section 2, it is possible that the growth of passive ownership has led investors to gather more information about systematic risk e.g., news about monetary policy. As a result, stock prices could become more informative about systematic information, at the expense of firm-specific news. Another possible effect of growing passive ownership is that fewer investors decide to become informed overall, and price informativeness decreased before the release of systematic information as well. I find that there is no significant downward trend in any of the three price informativeness measures around scheduled FOMC announcements. This confirms that the decrease in average 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 passive ownership and pre-earnings abnormal turnover, the pre-earnings drift and the share of volatility on earnings days. Finally, I calibrate the model to qualitatively match the cross-sectional results.

4.1 Pre-earnings turnover

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

$$CAT_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{22}$$

where cumulative abnormal pre-earnings turnover, $CAT_{i,t}$, is defined in Equation 19. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2 (i.e., returns used to form momentum portfolios in Jegadeesh and Titman (1993)), one-month lagged book-to-market ratio and total institutional ownership. $X_{i,t}$ also includes CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility, all computed over the previous 252 trading days. These controls are included because they capture many firm-specific characteristics known to be correlated with passive ownership (see e.g., Glosten et al. (2021)). $CAT_{i,t}$ is Winsorized at the 1% and 99% level each year.

Regression 22 also includes firm fixed effects, year/quarter fixed effects and quarter-of-the-year fixed effects. The year/quarter fixed effects, ϕ_t, ensure I am comparing firms at the same point in time with different levels of passive ownership, accounting for the time trends in price informativeness. The quarter-of-year fixed effects, ζ_q, account for seasonality. The firm fixed effects, ψ_i, account for firm-specific differences in average price informativeness e.g., investors pay more attention to Apple's earnings announcements than to those of Dominion Energy. Standard errors are double-clustered at the firm and year/quarter level.

Table 3 contains the regression results. In column 1, the right-hand-side only has passive ownership and the three sets of fixed effects. I find that passive ownership is negatively correlated with pre-earnings abnormal turnover. In column 2, the right-hand-side variables are the same as column 1, but I restrict to the sample of firm/quarter observations with non-missing control variables. The coefficient is almost unchanged, so the selection effect of restricting only observations that have non-missing control variables is not driving my results. Finally, in column 3, I add in all the firm-level controls in $X_{i,t}$. The coefficient on passive ownership shrinks, but is still economically large and statistically significant. I will refer to this specification, with equal weights, all controls and fixed effects as the baseline specification going forward.

The baseline specification (column 3) implies that a 15% increase in passive ownership would lead to a decline in cumulative abnormal pre-earnings turnover of -1.68. This effect is economically large, especially relative to the average decline in pre-earnings turnover of about 1 between the early 1990s and late 2010s.[13] To allay concerns that small firms are driving my results, columns 4 and 5 replicate columns 2 and 3, but within each quarter, firms are weighted by their last-quarter market capitalization. Using value weights, instead

[13]One might believe that because passive funds trade little, passive ownership mechanically decreases turnover. By focusing on *abnormal* turnover, these regression results suggest there is a decline in trading volume before earnings announcements *relative* to firm-level average turnover, allaying this concern.

of equal weights, does not lead to a statistically significant difference in the estimated effect of passive ownership on pre-earnings abnormal turnover.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-13.52***	-14.03***	-11.23***	-9.826***	-10.27***
	(2.947)	(2.965)	(2.958)	(3.092)	(2.852)
Observations	398,216	381,318	381,318	381,318	381,318
R-Squared	0.06	0.061	0.084	0.148	0.149
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table 3 Passive ownership and pre-earnings turnover. Estimates of β from:

$$CAT_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t}$$

where $CAT_{i,t}$ is cumulative abnormal pre-earnings turnover. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year/quarter level 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:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{23}$$

where $DM_{i,t}$ is defined as in Equation 20. All the controls and fixed effects are the same as in Equation 22. The regression results are in Table 4. The coefficient on $Passive_{i,t}$ in the baseline specification (column 3) implies that a 15% increase in passive ownership would decrease the pre-earnings drift by -0.0080. This effect is also economically large, at about 40% the size of the average decline in the drift over my sample of 0.02. The other columns show this result is not sensitive to the inclusion of firm-level controls, or using value weights.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.0432***	-0.0478***	-0.0528***	-0.0568***	-0.0474***
	(0.006)	(0.006)	(0.006)	(0.018)	(0.014)
Observations	447,069	417,109	417,109	417,109	417,109
R-Squared	0.197	0.201	0.217	0.249	0.266
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table 4 Passive ownership and pre-earnings drift. Table with estimates of β from:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t}$$

Where $DM_{i,t}$ is a measure of the pre-earnings drift. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year/quarter level in parenthesis.

4.3 Earnings days' share of volatility

I run the following regression to measure the relationship between earnings days' share of volatility, and passive ownership:

$$QVS_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{24}$$

where QVS is defined in Equation 21. All the controls and fixed effects are the same as in Equation 22. The regression results are in Table 5. The baseline specification (columns 3) implies that a 15% increase in passive ownership would lead to an increase in earnings-day volatility of 0.061. This is large, at about 1/3 the size of the average decline in QVS over the whole sample. This result is not significantly changed by including the firm-level controls, but it is weakened by about 50% when using value weights, instead of equal weights.

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

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	0.524***	0.501***	0.408***	0.214*	0.232**
	(0.027)	(0.028)	(0.031)	(0.111)	(0.094)
Observations	446,530	416,609	416,609	416,609	416,609
R-Squared	0.217	0.22	0.222	0.233	0.234
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table 5 Passive ownership and earnings days' share of volatility. Table with estimates of β from:

$$QVS_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t}$$

where $QVS_{i,t}$ is a measure of earnings days' share of volatility. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year/quarter level in parenthesis.

earnings news.

4.4 Robustness of cross-sectional regression results

One concern with the last set of cross-sectional regression results is that passive ownership increases overall volatility (see e.g., Ben-David et al. (2018)). Given the upward trend in passive ownership, this could create a spurious correlation between passive ownership and QVS. Recall that in Equation 21, the denominator contains volatility relatively further in the past than the numerator. If increases in passive ownership are persistent at the firm level, QVS might mechanically increase over time, as volatility further in the past would tend to be lower than current volatility. To rule out this possibility, and confirm that my results are specific to earnings days, I perform three placebo tests.

As in Section 3, the first set of placebo earnings dates are 22 trading days before each earnings announcement. The second are randomly selected dates each quarter. The third are all scheduled FOMC meetings. For the first two sets of placebo earnings announcements, there is no relationship between QVS and passive ownership. This suggests that my results are specific to earnings announcement dates. For the FOMC announcements, however, there is a weakly statistically significant positive relationship between passive ownership and QVS, but the magnitude is 1/20th as large as the coefficient in Table 5. This suggests that passive ownership may be correlated with decreased pre-FOMC announcement price informativeness, but the effect is quantitatively much smaller than for stock-specific news announcements.

Two additional threats to identification are (1) Regulation Fair Disclosure (Reg FD), passed in August 2000, 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 (2018), Farboodi and Veldkamp (2020)). The Online Appendix shows that all the cross-sectional results are robust to only using data after Reg FD passed i.e., earnings announcements from 2001-2018. The results are also robust to controlling for the AT measures in Weller (2018).[14] 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 around 15%. Recall that passive ownership is an equilibrium object, so I cannot set it directly. I target the level of passive ownership indirectly by setting the intermediary's risk aversion ρ^{int}. Second, price informativeness monotonically decreases after introducing the ETF in zero average supply and growing the ETF to around 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 i.e., the cost of becoming informed c (2) risk aversion ρ (3) the volatility of the systematic risk factor σ_f and (4) risk aversion of the ETF intermediary ρ^{int}.

[14]These measures are constructed from the SEC's MIDAS data, which starts in 2012. This lack of a long historical time series is why I do not include these as controls in my baseline cross-sectional regression specifications.

The results are in Table 6. Pre-earnings turnover 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, 30% learn when the ETF is 16% of the market, evidence of the extensive margin effect at work. Attention on systematic risk increases monotonically from 0.34 to 0.41, evidence that passive ownership also affects the intensive learning margin. In the Online Appendix, I perform a counterfactual analysis, examining the effect of increasing passive ownership to 50% of the market with these same parameters.

This calibration reveals the relative importance of the three main channels for explaining the decline in price informativeness. First, it is clear that the diversification and market timing channels dominate the hedging channel: fewer investors become informed and the remaining investors focus their attention on systematic risk. The effect of the diversification channel, however, is quantitatively more important than the market timing channel. To test this, I fix attention on systematic risk (possibly suboptimally) at 0.34 and directly vary the share of informed investors from 0.6 to 0.3. In this world, introducing and growing passive ownership leads all three price informativeness measures to decline about 90% as much as the *Change* row of Table 6. This suggests that the extensive margin effect i.e., the diversification channel explains most of the decrease in price informativeness.

The row labeled data is the effect of a 16% increase in passive ownership based on estimates from the baseline cross-sectional regressions. Although I am able to qualitatively mirror the empirical patterns with this calibration, the match is not always quantitatively strong. The match on volatility is relatively close, but the matches on volume and drift are off by factors of $4\times$ and $8\times$. One explanation for this is that in reality, there are many days between earnings announcements, while in the model, there is only one day, $t = 1$, to trade on earnings news before it is made public.

Although the model can qualitatively match the empirical results for the three measures of price informativeness, it does not imply that passive ownership *causes* a decrease in price informativeness. Inside the model, it's possible that a change in investors' preferences, and as a result, changes in learning, could lead the ETF intermediary to endogenously supply more shares of the ETF. Outside the model, it's possible that passive ownership endogenously increased the most in stocks that had the biggest decrease in price informativeness for unrelated reasons. In the next section, I exploit quasi-exogenous increases in passive

	ρ^{int}	ETF Size	Volume	Drift	Volatility	Share Informed	Attn. on Sys. Risk
Model	N/A	No ETF	0.969	0.9636	0.706	0.6	0.34
	9	0%	0.614	0.9628	0.783	0.35	0.35
	0.6	16%	0.505	0.9626	0.810	0.3	0.41
	Change		-0.464	-0.0010	0.104		
Data	16% Δ Passive		-1.80	-0.0084	0.0653		

Table 6 Model calibration to match cross-sectional regression results. The cost of becoming informed is set so 60% learn in equilibrium when the ETF is not present, $\rho = 0.15$, and σ_f=0.3. To roughly match the average passive ownership of 15% in 2018, ρ^{int} is set to 0.6, which leads to an equilibrium level of passive ownership in the model of 16%. Data row is effect of a 16% increase in passive ownership based on equal-weighted cross-sectional estimates with all controls and fixed effects.

ownership to rule out reverse causality.

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) Change 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

Four times a 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, being representative of the US economy 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 representativeness 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. To this end, at the time of index addition, I sort firms into three-digit SIC industries, and within each industry, I form quintiles of market capitalization. For each added firm, the first set of control firms are those in the same three-digit SIC industry and same quintile of industry market capitalization which are outside the S&P 500 index.

I also form a second control group of firms meeting the same selection criteria (3-digit SIC industry, industry market capitalization quintile), but are already in the S&P 500 index. This results in about 500 treated firms, 600 control firms out of the index, and 2,000 control firms in the index. Control firms can appear more than once e.g., the same firm in the index can be a control firm for multiple firms added to the index at different points in time. To identify the causal effect of passive ownership on stock price informativeness, I use index addition as an instrument for passive ownership. The first stage regression is:

$$Passive_{i,t} = \alpha + \beta_1 Post_{i,t} + \beta_2 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \tag{25}$$

Here, $Treated_{i,t}$ is equal to one if the firm was added to the S&P 500, and zero otherwise.[15] There are fixed effects for each interaction between firm and cohort group, where cohort group is defined by the combination of: (1) month of index addition (2) SIC-3 industry and (3) industry market capitalization quintile. $Treated_{i,t}$ is not included in Equation 25 because each firm can only be in one of the three treatment groups within each cohort group, so this term would be washed out by the fixed effects.

Figure 11 shows the level of passive ownership for the control firms and treated firms around the month of index addition. Within each cohort group, I demean passive ownership

[15]One concern with defining treatment as being *added* to the index, and not *staying* in the index, is that firms may change their index status during the period of study. The results are robust to requiring treated firms to be out of the index for the whole pre-treatment period, and in the index for the whole post-treatment period (and applying similar filters for both groups of control firms). This, however, is not my preferred specification, as whether or not a firm stays in/out of the index is endogenous, and future index status is not known at the time of index addition.

to make the effect comparable across cohorts and across time. All three groups of firms have similar average pre-addition *changes* in passive ownership, although the firms already in the index have a higher average *level* of passive ownership.

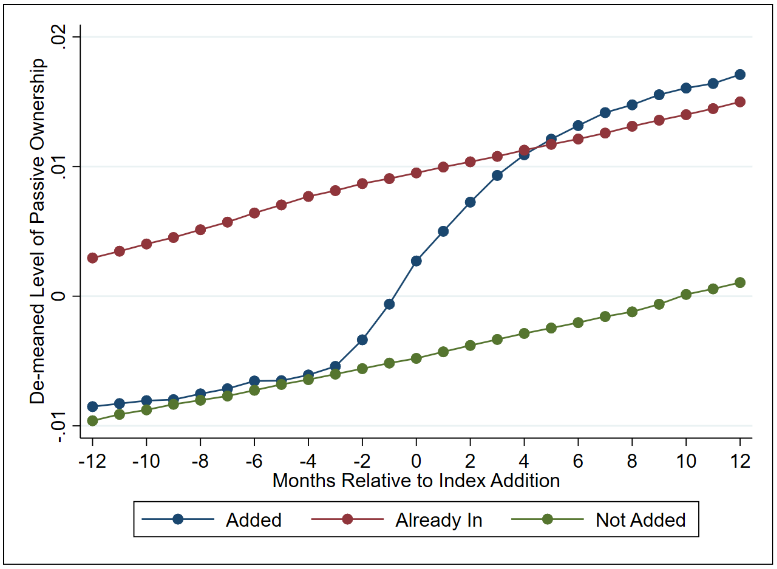

Figure 11. S&P 500 index addition and changes in passive ownership. Average level of passive ownership for control firms out of the index ("Not Added"), control firms in the index ("Already In") and treated firms ("Added"). Passive ownership is demeaned within each cohort group i.e., within each group of matched treated and control firms.

The three key pieces of my instrumental variables strategy are: (1) the instrumented change in passive ownership (2) the IV specification and (3) the reduced form specification:

$$\begin{aligned} \widehat{Passive}_{i,t} &= \alpha + \beta_1 Post_{i,t} + \beta_2 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \\ Outcome_{i,t} &= \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t} \\ Outcome_{i,t} &= \alpha + \beta_4 Post_{i,t} + \beta_5 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \end{aligned} \tag{26}$$

Where $Outcome_{i,t}$ is pre-earnings turnover, drift, or earnings days' share of volatility. The fixed effects are the same as in Equation 25. I restrict to data within three years of index addition, but exclude three months immediately before/after the month of index addition when computing these averages to: (1) avoid index inclusion effects (see e.g., Morck and Yang (2001)) and (2) avoid contamination by pre-trends in passive ownership starting at month $t = -2$ visible in Figure 11.

Because the change in passive ownership associated with being added to the S&P 500 has been increasing over time, I also run a specification that allows for heterogeneous treatment intensity:

$$\begin{aligned} Passive_{i,t} &= \alpha + \beta_1 Post_{i,t} + \beta_2 \text{Passive Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \\ Outcome_{i,t} &= \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t} \\ Outcome_{i,t} &= \alpha + \beta_4 Post_{i,t} + \beta_5 \text{Passive Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \end{aligned} \tag{27}$$

Here, Passive $\text{Gap}_{i,t}$ is the difference in passive ownership between the matched control firms in the index and out of the index, three months before index rebalancing. If at the time of index addition there are no matched control firms, I use the average Passive Gap for all firms that year. Passive $\text{Gap}_{i,t}$ is designed to capture the expected increase in passive ownership from being added to the index. The average Passive Gap at the end of my sample is about 4%.[16] The uninteracted term Passive $\text{Gap}_{i,t}$ is not included in Equation 27 because it is constant within each cohort group, and would be washed out by the fixed effects.

One concern with my research design 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.

Table 7 contains the regression results. Panel A examines the effect of index addition on pre-earnings turnover. Column 1 is the first stage regression. The associated F-statistic is huge, which is not surprising given the large increase in passive ownership pictured in Figure 11.[17] The coefficient on $Post \times Treated$ implies that the average increase in passive ownership associated with index addition is about 1.35%. Column 2 is the instrumental variables (IV) specification, where I use $Post$ and $Post \times Treated$ as instruments for passive ownership. The effect is negative and statistically significant, consistent with the cross-

[16]Due to the removal of S&P constituents from Compustat, the last cohort of index additions I consider is in 2017.

[17]Because I am using both $Post$ and $Post \times Treated$ as instruments for passive ownership, the time trend *and* the treatment effect in Figure 11 are driving the large magnitude of the F-statistic in Table 7. In a regression of passive ownership on $Post$, $Post \times Treated$, and the fixed effects, both terms are individually statistically significant, with the t-Statistic on $Post$ over 200 and the t-Statistic on $Post \times Treated$ over 40.

sectional regression results. Further, the IV estimate of -15.4 is not far from the baseline estimate of -11.2. Finally, column 3 is the reduced form regression, which provides an easy to interpret magnitude: Being added to the index is associated with a drop in pre-earnings abnormal turnover of -0.62.

Columns 4-6 repeat columns 1-3, but using the gap in passive ownership between the two sets of matched control firms, interacted with the treatment dummy, along with *Post* as instruments for passive ownership. As with the binary treatment, the first stage is economically large and statistically significant. The IV results are also similar to the baseline cross-sectional estimates. The reduced-form regression implies that a 4% larger Passive Gap i.e., expected increase in passive ownership (the average at the end of my sample) would lead to an average drop in pre-earnings abnormal turnover of about 1.

Panel B is the pre-earnings drift analogue of panel A. The first stage in panel B (and panel C) is nearly identical to panel A, as the only difference is in the set of firms that can be matched to non-missing drift data vs. non-missing turnover data. The IV estimate is the same sign as the cross-sectional coefficient of -0.05, but is about three times as large as the baseline. One possible reason for this is that my measure of passive ownership understates the true level of passive ownership firms experience after being added to the S&P 500 index.

In panel B, the reduced form estimate is not statistically significant in the continuous treatment specification. It is not obvious, however, that the reduced form regressions *should* be comparable with the baseline cross-sectional regression estimates. For the binary treatment specifications, the reduced form regression ignores the fact that the change in passive ownership from being added to the index increased from about 50bp in the early 1990s to 4% by the late 2010s. The continuous treatment specification partially addresses this issue, but given that both the model and the baseline regression estimates are about the *level* of passive ownership, it's not obvious why the expected *change* in passive ownership from index addition should be informative about anything other than the *sign* of the treatment effect.

Panel C examines earnings-day volatility. Both of the IV estimates are consistent with the cross-sectional estimates, but with coefficients about $3\times$ as large as the baseline specification. The reduced-form regressions are the right sign, but statistically insignificant.

Panel A: Pre-Earnings Volume	(1)	(2)	(3)	(4)	(5)	(6)
Post x Treated	0.0135***		-0.621***			
	(0.000)		(0.208)			
Passive Ownership		-15.41***			-15.57***	
		(3.240)			(3.209)	
Post x Treated x Passive Gap				0.550***		-24.90***
				(0.014)		(6.598)
Observations	188,606	188,606	188,606	188,606	188,606	188,606
F-statistic	33,189			33,120		

Panel B: Pre-Earnings Drift	(1)	(2)	(3)	(4)	(5)	(6)
Post x Treated	0.0135***		-0.00194**			
	(0.000)		(0.001)			
Passive Ownership		-0.172***			-0.168***	
		(0.013)			(0.013)	
Post x Treated x Passive Gap				0.550***		-0.028
				(0.014)		(0.034)
Observations	190,940	190,940	190,940	190,940	190,940	190,940
F-statistic	33,396			33,361		

Panel C: Earnings-Day Volatility	(1)	(2)	(3)	(4)	(5)	(6)
Post x Treated	0.0134***		0.00591			
	(0.000)		(0.005)			
Passive Ownership		1.513***			1.502***	
		(0.063)			(0.064)	
Post x Treated x Passive Gap				0.551***		0.195
				(0.014)		(0.206)
Observations	192,469	192,469	192,469	192,469	192,469	192,469
F-statistic	33,823			33,778		

Table 7 Effects of S&P 500 index addition on price informativeness. Estimates from:

$$\widehat{Passive}_{i,t} = \alpha + \beta_1 Post_{i,t} + \beta_2 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t}$$

$$Outcome_{i,t} = \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t}$$

$$Outcome_{i,t} = \alpha + \beta_4 Post_{i,t} + \beta_5 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t}$$

Columns 1 and 4 are first-stage regressions. Columns 2 and 5 are instrumental variables regressions. Columns 3 and 6 are reduced-form regressions. Columns 1-3 are from the binary treatment specifications, while columns 4-6 are from the continuous treatment specifications. FE are fixed effects for each interaction between firm and cohort group. Robust standard errors in parenthesis.

5.2 Russell 1000/2000 index rebalancing

The Russell 3000 contains approximately the 3000 largest stocks in the US stock market. At the end of each May, FTSE Russell selects the 1000 largest stocks by market capitalization to be members of the Russell 1000, while it selects the next 2000 largest stocks to be members of the Russell 2000. Both of these indices are value-weighted, so moving from the 1000 to the 2000 increases the fraction of a firm's shares owned by passive funds. This is because *switchers* go from being the smallest stock in an index of big stocks, to the biggest stock in an index of small stocks, significantly boosting their index weight (see e.g., Appel et al. (2016)).

To reduce turnover between the two indices, in 2007 Russell switched to a bandwidth rule, rather than using a sharp cutoff. As long as a potential switcher's market capitalization is within ± 2.5% of the Russell 3000E's total market capitalization, relative to the 1000th ranked stock, it will remain in the same index it was in the previous year.

The ideal experiment is to compare potential switchers to those that actually switched. This, however, is not straightforward, as the data that Russell uses to compute May market capitalizations is not made available to researchers. I follow the method in Coles et al. (2020) to compute a proxy for the Russell May market capitalizations.[18] Between 2007 and 2020, I am able to correctly predict Russell 1000/2000 index membership for 99.63% of Russell 3000 stocks overall, and 98.27% of Russell 3000 stocks within 100 ranks of the upper and lower bands.

I identify treated and control firms using the method in Coles et al. (2020). Each May I identify a cohort of possible switchers: those within +/- 100 ranks around the lower threshold that were in the Russell 1000 last year.[19] The treated firms are those that were over the threshold, and ended up switching, while the control firms are those that were above the

[18] I would like to thank the authors for sharing their replication code. I also incorporate the improvement discussed in Ben-David et al. (2019), which accounts for the exact day Russell rebalances its indices, rather than using end-of-month market capitalizations. The Online Appendix contains a step-by-step explanation of how I compute the May market capitalization proxy. For more details, see e.g., Chang et al. (2015), Wei and Young (2017), Gloßner (2018) and Heath et al. (2021).

[19] Another natural set of treated/control firms are those within 100 ranks of the upper band that were in the Russell 2000 the previous year; these are possible switchers to the Russell 1000, and they experience a decrease in passive ownership if they end up moving from the Russell 2000 to the Russell 1000. In the Online Appendix, I show that within one year of switching, the treatment effect is totally washed out by the time trend toward increased passive ownership.

threshold and stayed in the 1000.[20] A firm can be treated more than once if it switches to the 2000, goes back to the 1000, and then switches back to the 2000 at some future date. Control firms can appear more than once if they stay around the lower cutoff for multiple years, but never actually switch. These filters yield about 700 treated firms and 600 control firms.

Finally, I focus on Russell index reconstitutions after the rule change in 2007 for two reasons: (1) The average increase in passive ownership is larger than earlier years, as over time, more money has started to track these indices. Specifically, for switching firms, the *total* average increase in passive ownership each cohort (equal weighted from 2007 to 2018) is over 4%. The same average from 1990 to 2006 is around 1%. This ensures that my first stage regression will not be weak and (2) Under the bandwidth regime, switching is harder to predict/manipulate, so switching is less likely to be front-run.

Figure 12 compares the level of passive ownership around the index rebalancing date between the treated and control group. Within each cohort, I remove the mean level of passive ownership. Passive ownership starts increasing in the treated firms about two months 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.

For the Russell experiment, I use a setup similar to the S&P 500 experiment, with three key differences: (1) Passive $\text{Gap}_{i,t}$ is now defined as the difference in passive ownership between firms in the Russell 1000 and the Russell 2000 within $\pm$ 100 ranks around the cutoff in May, before index rebalancing (2) there are fixed effects for each interaction between firm and cohort group, but cohort group is now defined only by month of index rebalancing and (3) the time period is different, as I only use Russell reconstitutions between 2007 and 2018.

Table 8 contains the regression results. The first-stage results are positive, with a large F-statistic.[21] The estimated coefficient of 1.42%, however, understates the *total* change in passive ownership. In Figure 12, the average total increase for treated firms is around 4.5%, but there is about a 3% increase for the control firms, driven by the overall trend upward in

[20] As with the S&P experiment, the results are robust to requiring all of the treated/control firms to stay in their respective indices for the next 3 years. This, however, is endogenous, so I do not apply this filter in my baseline specification.

[21] As with the S&P experiment, this is driven both by the $Post$ and $Post \times Treated$ terms. In a regression of passive ownership on $Post$, $Post \times Treated$, and the fixed effects, both terms are individually statistically significant, with the t-Statistic on $Post$ over 50 and the t-Statistic on $Post \times Treated$ over 9.

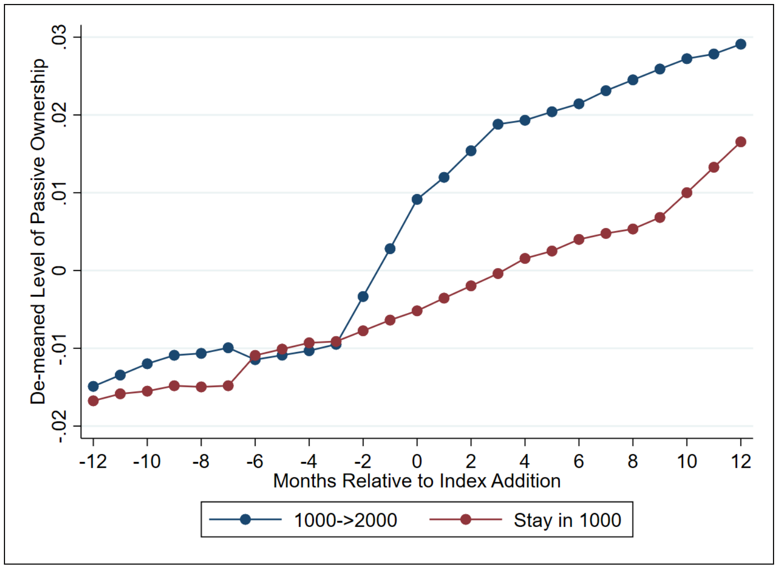

Figure 12. Russell 1000/2000 reconstitution and changes in passive ownership. Average level of passive ownership for firms that stay in the Russell 1000 (control firms) and firms that moved from the Russell 1000 to the Russell 2000 (treated firms). Passive ownership is demeaned within each cohort.

passive ownership.

All the estimated coefficients are qualitatively consistent with the cross-sectional regression estimates: switching to the Russell 2000 is associated with a drop in pre-earnings turnover, a drop in the pre-earnings drift, and an increase in earnings-day volatility. The IV estimates of -23.09, -0.11 for the volume and drift specifications are significantly larger than the estimates from Tables 3 and 4, while the IV estimate of 0.23 for volatility is similar in magnitude to those in Table 5.

As with the S&P experiment, some of the reduced-form regressions are statistically insignificant. Again, this is not surprising as these reduced-form specifications do not account for (1) the time-series increase in the treatment effect even within the 2007 to 2018 sample and (2) differences in the resulting *level* of passive ownership across cohorts.

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, passive ownership is correlated with decreased price informativeness. Stocks with more passive ownership have less pre-earnings turnover, smaller pre-earnings drifts and a larger share of volatility on earnings days. To rule out reverse causality, 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 small, owning only 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.

	(1)	(2)	(3)	(4)	(5)	(6)
	Panel A: Pre-Earnings Volume					
Post x Treated	0.0142***		-0.742			
	(0.002)		(0.503)			
Passive Ownership		-22.89***			-23.09***	
		(4.690)			(4.717)	
Post x Treated x Passive Gap				0.451***		-51.93**
				(0.087)		(24.320)
Observations	5,893	5,893	5,893	5,893	5,893	5,893
F-statistic	2,533			2,542		
	Panel B: Pre-Earnings Drift					
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Treated	0.0139***		-0.00615**			
	(0.002)		(0.003)			
Passive Ownership		-0.116***			-0.113***	
		(0.027)			(0.027)	
Post x Treated x Passive Gap				0.441***		-0.248*
				(0.084)		(0.138)
Observations	6,174	6,174	6,174	6,174	6,174	6,174
F-statistic	2,677			2,686		
	Panel C: Earnings-Day Volatility					
	(1)	(2)	(3)	(4)	(5)	(6)
Post x Treated	0.0140***		0.000894			
	(0.002)		(0.014)			
Passive Ownership		0.216*			0.230*	
		(0.127)			(0.128)	
Post x Treated x Passive Gap				0.452***		0.741
				(0.084)		(0.659)
Observations	6,175	6,175	6,175	6,175	6,175	6,175
F-statistic	2,675			2,687		

Table 8 Effects of Russell 1000/2000 index reconstitution on price informativeness. Estimates from:

$$\widehat{Passive}_{i,t} = \alpha + \beta_1 Post_{i,t} + \beta_2 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t}$$

$$Outcome_{i,t} = \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t}$$

$$Outcome_{i,t} = \alpha + \beta_4 Post_{i,t} + \beta_5 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t}$$

Columns 1 and 4 are first-stage regressions. Columns 2 and 5 are instrumental variables regressions. Columns 3 and 6 are reduced-form regressions. Columns 1-3 are from the binary treatment specifications, while columns 4-6 are from the continuous treatment specifications. FE are fixed effects for each interaction between firm and cohort group. Robust standard errors in parenthesis.

References

Admati, A. R. (1985). A Noisy Rational Expectations Equilibrium for Multi-Asset Securities Markets. *Econometrica*, 53(3).

Admati, A. R. and Pfleiderer, P. (1988). A Theory of Intraday Patterns: Volume and Price Variability. *Review of Financial Studies*, 1(1).

Akerlof, G. A. (1970). The market for “lemons”: Quality uncertainty and the market mechanism.

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

Bai, J., Philippon, T., and Savov, A. (2016). Have financial markets become more informative? *Journal of Financial Economics*, 122(3).

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*, 6(2).

Ben-David, I., Franzoni, F., and Moussawi, R. (2018). Do ETFs Increase Volatility? *Journal of Finance*, 73(6).

Ben-David, I., Franzoni, F. A., and Moussawi, R. (2019). An Improved Method to Predict Assignment of Stocks into Russell Indexes. *SSRN Electronic Journal*.

Beschwitz, B. v., Honkanen, P., and Schmidt, D. (2020). Passive ownership and short selling. *Available at SSRN*.

Bhattacharya, A. and O’Hara, M. (2018). Can ETFs Increase Market Fragility? Effect of Information Linkages in ETF Markets. *SSRN Electronic Journal*.

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

Buss, A. and Sundaresan, S. (2020). More Risk, More Information: How Passive Ownership Can Improve Informational Efficiency. *SSRN Electronic Journal*.

Campbell, J. Y., Lettau, M., Malkiel, B. G., and Xu, Y. (2001). Have individual stocks become more volatile? An empirical exploration of idiosyncratic risk. *Journal of Finance*, 56(1).

Chang, Y. C., Hong, H., and Liskovich, I. (2015). Regression discontinuity and the price effects of stock market indexing. *Review of Financial Studies*, 28(1).

Chinco, A. and Fos, V. (2021). The Sound of Many Funds Rebalancing. *The Review of Asset Pricing Studies*.

Coffee, J. C. (2007). Law and the market: The impact of enforcement.

Coles, J. L., Heath, D., and Ringgenberg, M. (2020). On Index Investing. *SSRN Electronic Journal*.

Cong, L. W., Huang, S., and Xu, D. (2020). Rise of factor investing: security design and asset pricing implications. Technical report, Working Paper.

Dávila, E. and Parlatore, C. (2018). Identifying Price Informativeness. *National Bureau of Economic Research Working Paper Series*.

Dávila, E. and Parlatore, C. (2021). Trading Costs and Informational Efficiency. *Journal of Finance*, 76(3).

Epstein, L. G. and Zin, S. E. (1989). Substitution, Risk Aversion, and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework. *Econometrica*, 57(4).

Ernst, T. (2020). Stock-Specific Price Discovery From ETFs. *Working Paper*, (October 2017).

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).

Farboodi, M., Matray, A., Veldkamp, L., and Venkateswaran, V. (2020). Where has all the data gone? Technical report, National Bureau of Economic Research.

Farboodi, M. and Veldkamp, L. (2020). Long-Run growth of financial data technology. *American Economic Review*, 110(8).

Foster, G., Olsen, C., and Shevlin, T. (1984). Earnings Releases, Anomalies, and the Behavior of Security Returns. *Source: The Accounting Review*, 59(4).

Garleanu, N. B. and Pedersen, L. H. (2018). Active and Passive Investing. *SSRN Electronic Journal*.

Gloßner, S. (2018). The Effects of Institutional Investors on Firm Outcomes: Empirical Pitfalls of Quasi-Experiments Using Russell 1000/2000 Index Reconstitutions. *SSRN Electronic Journal*.

Glosten, L., Nallareddy, S., and Zou, Y. (2021). ETF activity and informational efficiency of underlying securities. *Management Science*, 67(1).

Gorton, G. B. and Pennacchi, G. G. (1993). Security Baskets and Index-Linked Securities. *The Journal of Business*, 66(1).

Grossman, S. J. and Stiglitz, J. E. (1980). On the Impossibility of Informationally Efficient Markets On the Impossibility of Informationally Efficient Markets. *Source: The American Economic Review*, 70(3).

Haddad, V., Huebner, P., and Loualiche, E. (2021). How Competitive is the Stock Market? Theory, Evidence from Portfolios, and Implications for the Rise of Passive Investing. *SSRN Electronic Journal*.

Heath, D., Macciocchi, D., Michaely, R., and Ringgenberg, M. C. (2021). Do Index Funds Monitor? *The Review of Financial Studies*.

Jegadeesh, N. and Titman, S. (1993). Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency. *The Journal of Finance*, 48(1).

Kacperczyk, M., Van Nieuwerburgh, S., and Veldkamp, L. (2016). A Rational Theory of Mutual Funds' Attention Allocation. *Econometrica*, 84(2).

Kacperczyk, M. T., Nosal, J. B., and Sundaresan, S. (2018). Market Power and Price Informativeness. *SSRN Electronic Journal*.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. *Econometrica*, 53(6).

Lee, J. (2020). Passive Investing and Price Efficiency. *SSRN Electronic Journal*.

Malikov, G. (2020). Information, Participation, and Passive Investing. *SSRN Electronic Journal*.

Martineau, C. (2018). The Evolution of Market Price Efficiency Around Earnings News. *SSRN Electronic Journal*.

Mauboussin, M. (2019). Looking for Easy Games in Bonds. *Bluemountain Investment Research*.

Mclean, R. D. and Pontiff, J. (2016). Does Academic Research Destroy Stock Return Predictability? *Journal of Finance*, 71(1).

Merton, R. C. (1987). A Simple Model of Capital Market Equilibrium with Incomplete Information. *The Journal of Finance*, 42(3).

Milgrom, P. and Stokey, N. (1982). Information, trade and common knowledge. *Journal of Economic Theory*, 26(1).

Morck, R. and Yang, F. (2001). The mysterious growing value of s&p 500 membership. Technical report, National Bureau of Economic Research.

Novy-Marx, R. (2015). Fundamentally, Momentum is Fundamental Momentum. *National Bureau of Economic Research*.

Schipper, K. (1989). Commentary on earning management. *Accounting Horizons*, 3(4).

S&P, D. J. I. (2017). S&p u.s. indices methodology. Technical report.

Subrahmanyam, A. (1991). A Theory of Trading in Stock Index Futures. *Review of Financial Studies*, 4(1).

Veldkamp, L. L. (2011). *Information choice in macroeconomics and finance*.

Wang, J. (1994). A Model of Competitive Stock Trading Volume. *Journal of Political Economy*, 102(1).

Wei, W. and Young, A. (2017). Selection Bias or Treatment Effect? A Re-Examination of Russell 1000/2000 Index Reconstitution. *SSRN Electronic Journal*.

Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *Review of Financial Studies*, 31(6).