Passive Ownership and Price Informativeness

MARCO SAMMON *

February, 2022

ABSTRACT

How does passive ownership affect the incorporation of information into stock prices? Motivated by two canonical models, I propose three new empirical measures of price informativeness. 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. Finally, I provide evidence for a mechanism: investors expend less effort gathering information about stocks with a larger fraction of passive owners.

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

1   Introduction

The rise of passive ownership is one of the most significant changes in asset markets over the past 30 years. Passive funds grew from owning less than 1% of the US stock market in the early 1990s to owing nearly 15% in 2018. As passive ownership continues to grow it is increasingly important to understand: 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. 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 informed investors to trade less aggressively on their private 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. One reason for the literature's lack of consensus is that researchers do not agree on how to measure price informativeness. A model is typically needed to guide its measurement, but it is not straightforward to take model-specific measures of price informativeness to the data. For example, consider Grossman and Stiglitz (1980), where price informativeness is the conditional variance of fundamentals, given prices. This is often estimated empirically with a regression of future $(t + 1)$ fundamentals on current (t) prices.

The correct measure of future fundamentals, however, 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. This problem is not unique to Grossman-Stiglitz. In Kyle (1985), price informativeness is the conditional volatility of fundamentals given the sequence of order flows. This suffers from the same issue as before, relying on the unobserved fundamental value of the firm.

Earnings are typically used to measure fundamentals (see e.g., Bai et al. (2016)), but it is not clear that they are ideal for this because: (1) management has some control over earnings growth (see e.g., Schipper (1989)) and (2) firms are increasingly composed of growth options, which may not show up in short- or even medium-run earnings. Finally, the right

set of conditioning variables is not clear: An econometrician does not know which variables investors use along with the price when forming expectations.

I address these issues in two ways. First, rather than using a model-specific definition of price informativeness, I propose measures that arise in multiple classic models of trading under asymmetric information. Second, I use discrete information-release events to gain statistical power when mapping the models' predictions to the data.

I start with the intuition that price informativeness should increase if (1) informed investors receive more precise signals or (2) there is an increase in the share of informed investors. I vary these parameters in two canonical models – Kyle (1985) and Grossman and Stiglitz (1980) – focusing on their predictions for quantities easily measured in the data: trading volume, returns and volatility.

In Kyle (1985), volume is related to the precision of the informed investor's signal because the more precise that signal is, the more aggressively they bet on it. This increased aggressiveness leads to more trading. Returns are also related to the informed investor's signal precision because the higher this precision, the more sensitive prices will be to order flow. This means that, on average, prices will end up closer to the fundamental value of the asset before uncertainty is resolved, as more information is traded into prices. Finally, the combination of these two effects means that as the informed investor's signal becomes more precise, a larger share of total volatility will happen on the intermediate trading dates, relative to when uncertainty is resolved.

In Grossman and Stiglitz (1980), the mechanism is similar: informed investors will be more aggressive if they get more precise signals, which leads to more trading volume. There is, however, an additional extensive margin effect. As fewer investors become informed, there are fewer differences of opinion. All uninformed investors receive the same signal from the price and thus they all submit the same demand. In the limit, if all investors are uninformed, there is no trade. The argument for smaller absolute returns on the uncertainty resolution date is essentially identical to Kyle (1985).

Neither of these models, however, was designed to explicitly feature passive ownership. In the Online Appendix, I develop an extension of Admati (1985) with endogenous learning and an intermediary who can buy shares of the underlying stocks and put them into an ETF. This richer model reveals channels through which passive ownership affects investors' learning behavior and thus price informativeness.

3

Using the models' predictions for trading volume, returns and volatility, I define three new measures of price informativeness related to discrete information-release events: (1) pre-earnings volume, (2) pre-earnings drift and (3) earnings-day volatility. Through simulations, I show that as the precision of informed investors' signals decreases, or the share of informed investors decreases, pre-earnings trading volume declines, the pre-earnings drift declines and earnings-day volatility increases. I create empirical analogues of these measures and show that pre-earnings price informativeness 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 relatively 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 e.g., Coffee (2007)), all the cross-sectional regressions include year-quarter fixed-effects. These cross-sectional results imply that passive ownership decreases price informativeness.

Of course, 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.

Finally, motivated by the richer model, I provide evidence on a learning-based mechanism for why passive ownership affects price informativeness. In the cross-section, passive ownership is correlated with (1) decreased coverage and accuracy by sell-side analysts and (2) fewer downloads of SEC filings. I rule out several alternative explanations, showing that (1) all the trends and cross-sectional regression results are specific to earnings announcement dates (2) the results are robust to using an *ex-ante* measure of uncertainty and (3) the results are not driven by firms with more passive ownership having earnings news of a meaningfully different nature than firms with less passive ownership.

My paper contributes to two broad strands of literature. First, I contribute to the

4

literature on measuring and quantifying trends in price informativeness. Several recent papers have argued that average price informativeness has trended up over time (See e.g., Bai et al. (2016), Dávila and Parlatore (2018) and Farboodi et al. (2020)). I offer an alternative strategy for measuring price informativeness that exploits the timing of when information is released and does not rely on knowing the fundamental value of a firm.

With my new measures and empirical focus on discrete information-release events, I come to the opposite conclusion: average price informativeness has decreased over the past 30 years. Although these findings are contradictory, I am making a narrower claim. My results are about how well prices reflect information contained in earnings announcements just before that information is released to the public.

The second main literature that I contribute to is the ongoing debate on the relationship between passive ownership and price informativeness. Several papers argue that the introduction of ETFs and rising passive ownership can improve price informativeness by changing a firm's cost of capital, increasing liquidity and making short-selling easier (see e.g., Buss and Sundaresan (2020), Ernst (2020), Lee (2020), Malikov (2020), Beschwitz et al. (2020)). Other papers argue that passive ownership reduces price informativeness by decreasing investors' incentives to gather or trade on private information (see e.g., Garleanu and Pedersen (2018), Kacperczyk et al. (2018), Breugem and Buss (2019), Haddad et al. (2021)). Finally, in some papers, the effect is ambiguous or neutral because ETFs increase the incorporation of factor-specific information at the expense of stock specific information or increases in passive ownership are offset by increases in the number of informed investors (see e.g., Cong et al. (2020), Bhattacharya and O'Hara (2018), Coles et al. (2020), Glosten et al. (2021)).

I contribute to this debate in two ways. First, I am able to examine the relationship between passive ownership and price informativeness without relying on the assumptions of any particular model. This is because my measures of price informativeness are based on quantities which are directly observable in the data. Second, my empirical strategy exploits the exact timing of *when* information is released. Through this, I am able to provide a more precise answer of how and when information is incorporated into stock prices.

Paper Outline. Section 2 motivates three new measures of price informativeness which arise in two canonical models of trade under asymmetric information. 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. Finally, Section 6 presents evidence that passive ownership is correlated with less learning by investors and rules out several alternative explanations.

2   Measuring price informativeness

The goal of this section is to define measures of price informativeness which are straightforward to estimate empirically. Regardless of the exact modeling assumptions, these measures are designed to capture the following intuition: price informativeness should be higher if, all else equal, the share of informed investors increases and/or the precision of informed investors' signals increases. I define three measures that satisfy these criteria in Kyle (1985) and Grossman and Stiglitz (1980): pre-earnings trading volume, pre-earnings drift and earnings-days' share of volatility.

These models were not designed to capture the rise of index investing. I summarize the details of a richer model which explicitly accounts for the role of passive ownership. This model reveals channels through which passive ownership affects investors' learning behavior and thus price informativeness.

2.1   Kyle (1985)

Consider a Kyle (1985)-style model with two trading periods, $t = 1$ and $t = 2$.[1] There is a *strategic* risk-neutral informed investor who receives a signal at time 1, $s = v + \epsilon$, where v is the true value of the asset and ϵ is signal noise. v and ϵ are independent and normally distributed both with mean zero and standard deviations σ_v and σ_ϵ.

The informed investor submits demands to a *competitive* risk-neutral market maker at times 1 and 2, y_1 and y_2. To prevent prices from being fully revealing, there are a group of noise traders who submit random demands z_1 and z_2, where the z_t are independent and

---

[1]The model setup in this section borrows heavily from Alex Chinco's "Two Period Kyle (1985) Model" notes. The Online Appendix contains details on how the model is solved.

normally distributed with mean zero and standard deviation σ_z. The market maker only observes total order flow, $x_t = y_t + z_t$, and uses the sequence of x_ts to set prices.

In this economy, there will be three prices: p_1 and p_2, which are the prices in the first and second trading periods and $p_3 = v$. When mapping the model to data, I view $t = 1$ and $t = 2$ as dates before uncertainty is resolved i.e., pre-earnings announcement trading dates. I view $t = 3$ as the date where the fundamental value of the firm is revealed i.e., an earnings announcement.

2.2   Grossman and Stiglitz (1980)

Consider a Grossman and Stiglitz (1980)-style model with a unit mass of atomistic rational investors, ϕ of whom are informed and 1-ϕ who are uninformed.[2] There are also noise traders who submit random demand x which prevents prices from fully revealing the informed investors' signals. At $t = 2$, informed investor j gets noisy signal $s_j = v + \epsilon_j$. The key differences from Kyle (1985) are that: (1) there are *competitive* informed investors who each get a *different* signal, so prices aggregate diverse information (2) all investors are risk averse with coefficient of absolute risk aversion ρ and (3) instead of a market maker who learns about v from order flow, there are uninformed investors who learn about v from prices.

Informed investor j updates their beliefs about v using the price p along with their signal s_j and submits a demand function q_j. All uninformed investors update their beliefs about v using only the price p and submit the same demand function q_u. There are initially $\overline{x}$ shares of the stock outstanding, so for market clearing to hold, it must be that $(1-\phi) \cdot q_u + \int_0^\phi q_j dj = \overline{x} + x$. There are two prices in the model, p_2 i.e., the price after all investors submit their demand functions and markets clear and $p_3 = v$ i.e., the fundamental value of the asset. In mapping the model to the data, I view $t = 2$ as the pre-earnings announcement period and $t = 3$ as the earnings announcement.

Although it is somewhat outside the scope of this model, as there is only one trading period, I would also like to compute p_1 i.e., the price before investors become informed and submit their demands. Suppose investors had to pay a fixed cost to become informed. Then in equilibrium, the expected utility of paying the cost and becoming informed would have to

---

[2]The model setup in this section borrows heavily from Admati (1985). Details on how the model is solved are in the Online Appendix.

be equal to the expected utility of remaining uninformed. One way to view this equilibrium is that every investor plays a mixed strategy where they become informed ϕ percent of the time and stay uninformed $1 - \phi$ percent of the time. p_1 is the price such that investors would be happy to hold their endowment given the strategy they are going to play and the strategies everyone else is going to play *before* they find out whether they will be informed or not. Assuming there is no time discounting and no risk aversion over whether they will be informed or uninformed, I set $p_1 = E[p_2]$.

2.3 Measures of price informativeness

Pre-Earnings Trading Volume

In Kyle (1985), the informed investor follows a linear demand rule: $y_t = \alpha_{t-1} + \beta_{t-1} \cdot s$, where α_{t-1} and β_{t-1} are constants that depend on model parameters. $\beta_t > 0$ so better signals lead to larger orders. With this in mind, the first price informativeness measure, pre-earnings trading volume, is defined as: $Volume = \sum_{t=1}^{2} |y_t + z_t|$

Hypothesis 1: Higher $Volume$ implies higher price informativeness

The intuition is that for a given signal size, $|s|$, orders will be larger if informed investor aggressiveness, β_t, is higher. Aggressiveness is proportional to $\sigma_v/(\sigma_v + \sigma_\epsilon)$ i.e., it is increasing in the precision of the informed investor's signal. Therefore, higher values of $Volume$ imply more precise signals and thus higher price informativeness.

The definition of $Volume$ is slightly different in Grossman and Stiglitz (1980). Suppose each investor starts with an equal share of the overall endowment. Then define pre-earnings volume as the difference between each rational investor's initial position and realized demands: $Volume = (1 - \phi) \cdot (|q_u - \overline{x}|) + \int_0^\phi |q_j - \overline{x}| dj$.

Two factors that drive trading volume in Grossman and Stiglitz (1980) are: (1) The share of investors who are informed. As more investors become informed, there are more investors with different posterior beliefs and thus more trading. Uninformed investors all submit the same demand because they all use the same signal from prices to form their posterior beliefs. All investors have the same initial endowment, so if there were only uninformed investors, there would be no trading volume. (2) Signal precision. As signals become more precise, informed investors have more precise posterior beliefs and are more willing to bet aggressively on their signals. Less trading volume is therefore evidence of fewer informed investors and

less precise signals.

Using trading volume to quantify price informativeness is motivated not only by these models, 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) and Wang (1994)).

Pre-Earnings Drift

In Grossman and Stiglitz (1980), define the cumulative return from period a to b as $r_{(a,b)} = (p_b - p_a)/p_a$. Define the pre-earnings drift as:

$$DM = \begin{cases} \frac{1+r_{(1,2)}}{1+r_{(1,3)}} & \text{if } r_{(2,3)} > 0 \\ \frac{1+r_{(1,3)}}{1+r_{(1,2)}} & \text{if } r_{(2,3)} < 0 \end{cases} \Leftrightarrow \begin{cases} \frac{1}{1+r_{(2,3)}} & \text{if } r_{(2,3)} > 0 \\ 1+r_{(2,3)} & \text{if } r_{(2,3)} < 0 \end{cases} \tag{1}$$

In words, DM is the fraction of the total gross return, $1 + r_{(1,3)}$, that occurs before uncertainty is resolved. Intuitively, values of DM closer to 1 suggest more informative pre-earnings announcement prices. For example: Suppose a stock appreciates 4% in the month leading up to an earnings announcement and increases 1% on the earnings announcement day itself. Compare this to the case where a stock appreciates 1% in the month prior to the announcement, but jumps 4% on the day-of. In the first case, it seems as though more of the earnings information was traded into prices before it was formally released. If $r_{(2,3)}$ is negative, this relationship would be reversed, which is why the measure is inverted when $r_{(2,3)}$ is less than zero.

Hypothesis 2: Higher DM implies higher price informativeness

If there are more informed traders or informed investors receive more precise signals, their trading will push $t = 1$ prices closer to fundamentals.

In Kyle (1985), define the return from period a to b as $r_{(a,b)} = p_b - p_a$.[3] Given the alternative formulation for DM in Equation 1 that depends only on $r_{(2,3)}$, define the pre-earnings drift as: $DM = 1 - |r_{(2,3)}|$. DM is related to the precision of the informed investor's

---

[3]Unlike in Grossman and Stiglitz (1980), both the informed investor and the market maker are risk neutral, so prices can be negative. This is why I define returns as level changes in the price, rather than percentage changes in the price.

9

signals because of the market maker's linear pricing rule: $p_t = \kappa_{t-1} + \lambda_{t-1} \cdot x_t$, where κ_t and λ_t are constants that depend on model parameters. λ_t is always positive, so larger total order flow leads to higher prices. Further, λ_t is *decreasing* in σ_ϵ, so when the informed investor has less information, prices respond less to order flow. This makes prices less sensitive to fundamental information and p_2 will on average be further from v, lowering DM.

Outside of these models, 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) and Weller (2018)). These papers argue that if investors are trading on signals of good news before earnings are released, or the firm gives guidance of strong future performance, we should expect prices to increase before the earnings announcement date.

Share of Volatility on Earnings Days

In Grossman and Stiglitz (1980), the share of total volatility on earnings days is: $\frac{r^2_{(2,3)}}{r^2_{(1,2)} + r^2_{(2,3)}}$. If prices are not informative before earnings announcements, we would expect earnings-day volatility to be large, relative to total volatility. This leads to my third price informativeness measure, QVS, defined as:

$$QVS = 1 - \frac{r^2_{(2,3)}}{r^2_{(1,2)} + r^2_{(2,3)}} \tag{2}$$

Hypothesis 3: Higher values of QVS imply higher price informativeness

As more investors become informed and as informed investors' signals become more precise, we expect a relatively larger difference between p_1 and p_2 and a relatively smaller difference between p_2 and p_3. I define QVS this way, rather than $\widehat{QVS} = \frac{r^2_{(2,3)}}{r^2_{(1,2)} + r^2_{(2,3)}}$, so that consistent with the other two measures, higher values of QVS imply more informative prices. While QVS is related to DM, the simulation results reveal how and why they can contain different information.

The definition of QVS is slightly different in Kyle (1985): $QVS = 1 - \frac{|r_{(2,3)}|}{|r_{(1,2)}| + |r_{(2,3)}|}$. Here, QVS is the fraction of the total distance traveled between $t = 1$ and $t = 3$ that occurs before uncertainty is resolved. The intuition for its relationship to the informed investor's signal precision, however, is identical to Grossman and Stiglitz (1980).

2.4 Comparative Statics

Effect of signal precision and the share of informed investors

For each set of parameters, I simulate the economy 10,000 times and compute averages of $Volume$, DM and QVS. The top panel of Figure 1 shows the relationship between the three price informativeness measures and the volatility of the informed investor's signal, σ_ϵ, in Kyle (1985). Consistent with the intuition outlined above, $Volume$, DM and QVS are all monotonically decreasing in σ_ϵ. This monotonic relationship, however, should not be viewed as a formal proof. It is possible to break the monotonic relationship between $Volume$ and the informed investor's precision in an economy with sufficiently low σ_v, low σ_z and high σ_ϵ.[4] For DM and QVS, the relationship is monotonic across a broad set of possible parameter choices.

The middle panel of Figure 1 shows the intensive information margin in the Grossman and Stiglitz (1980)-style model. Consistent with the top panel, $Volume$, DM and QVS are decreasing in σ_ϵ. The bottom panel shows the extensive information margin. As expected, $Volume$, DM and QVS are decreasing in the share of uninformed investors, $1 - \phi$. Again, this should not be viewed as a formal proof. While the relationship between the intensive and extensive learning margins and DM is monotonic across a broad set of possible parameter choices, it is possible to break the monotonicity in $Volume$ and QVS. As in Kyle (1985), this can happen when the informed investors' signals become sufficiently imprecise, or the share of uninformed investors becomes sufficiently large.

Effect of varying noise trader intensity

As discussed in Kyle (1985), changing *only* the volatility of noise trader shocks should not affect the conditional volatility of fundamentals, given the sequence of order flows. This is because noise increases market depth i.e., decreases λ_t, which encourages the informed investor to trade more aggressively i.e., increases β_t. The offsetting effect of these two forces is why the blue circles and red triangles, as well as the green diamonds and orange x's, are nearly overlapping in the top panel of Figure 1. It also explains why average abnormal

---

[4]There are two competing forces that govern the informed investor's trading behavior: (1) their desire to trade and (2) how costly it is to implement the desired trades. In an economy with low σ_v, low σ_z and low σ_ϵ, trading is initially costly: there is little fundamental uncertainty and the market maker knows they will mostly be trading with informed investors who have precise signals. As σ_ϵ increases, however, trading becomes cheaper because λ is decreasing in σ_ϵ. At some point, the cheaper trading force starts to dominate the desire to trade force and overall trading volume increases.
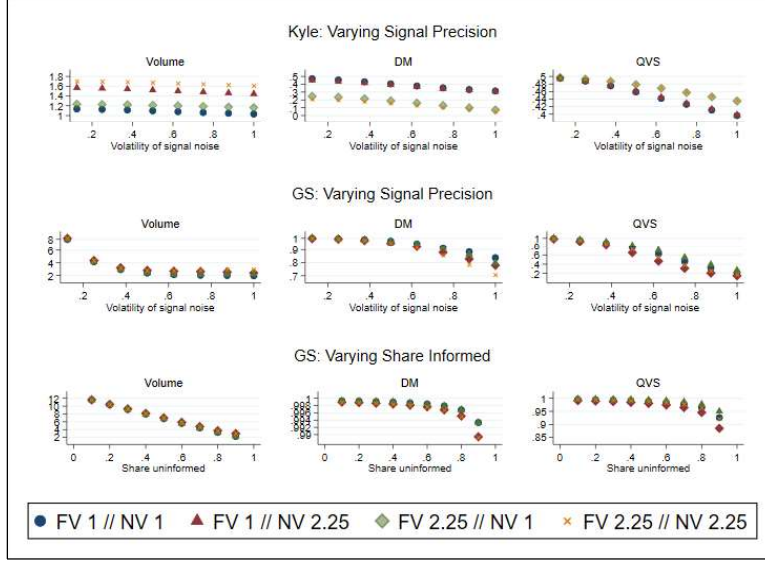
Figure 1. Informed investors' precision, the share of informed investors and price informativeness. Each point represents the average of the price informativeness measures across 10,000 simulations. FV=fundamental volatility=σ_v^2. NV=noise volatility=σ_z^2. In the second panel, I set the share of informed investors $\phi = 0.6$. In the bottom panel, I set $\sigma_\epsilon = 0.125$. In the bottom two panels , I set rational investors' risk aversion $\rho = 0.5$.

volume is monotonically increasing in the volatility of the noise shocks: informed investors trade more when it is easier to profit from their private information.

Differences between DM and QVS

The top panel of Figure 1 shows that QVS is relatively more responsive to changes in σ_ϵ when fundamental volatility is low, while DM is relatively more responsive to changes in σ_ϵ when fundamental volatility is high. As σ_z increases, the relative difference between $|r_{(2,3)}|$ and $|r_{(1,2)}|$ decreases. This is because prices become more sensitive to order flow, which leads to larger average deviations from the ex-ante expected price in the intermediate trading periods, increasing average QVS. This relative increase in $|r_{(1,2)}|$, however, has no effect on DM.

A similar argument explains why QVS is more sensitive to changes in σ_ϵ when fundamental volatility is low. Because ex-ante uncertainty is low, $|r_{(2,3)}|$ is smaller on average, which makes DM larger, regardless of the informed investor's signal precision. QVS, however, is still sensitive to σ_ϵ because the smaller values of $|r_{(2,3)}|$ are compared to the also smaller values of $|r_{(1,2)}|$. These different sensitivities to fundamental volatility suggest that

leveraging both measures is useful as they are cross-checks against one another.

Despite their different sensitivity to various parameters, it is not obvious that QVS and DM contain different information. Mechanically, QVS is not a function of DM because DM does not depend on $r_{(1,2)}$, or in a model with more periods, the returns in any period where uncertainty has not been totally resolved. A straightforward test for overlap is to run a regression of QVS on DM *within* each set of parameter choices, across simulations. I find these regressions have R-squared values of around 0.4. The differences between DM and QVS are driven by cases where DM is high but QVS is low e.g., $r_{(1,2)} = -1\%$ and $r_{(2,3)} = -1\%$. In this scenario, the $r_{(2,3)}$ return is relatively small, but volatility was equal in the intermediate trading period to when all uncertainty was resolved.[5]

2.5 Effects of passive ownership on price informativeness

Neither Kyle (1985) nor Grossman and Stiglitz (1980) was designed to model the effect of passive ownership on price informativeness. In the Online Appendix, I develop a richer model to explore the ways passive ownership affects investors' learning behavior. I start by introducing a new investor into an Admati (1985)-style model: an ETF intermediary who can buy shares of the underlying stocks and put them into an ETF. The model also features two key learning trade-offs: (1) a decision to pay a cost and become informed or stay uninformed i.e., the extensive learning margin and (2) whether to learn about systematic or idiosyncratic risk i.e., the intensive learning margin. The introduction of and increased importance of the ETF will affect both of these learning trade-offs.

There are two competing forces that govern the effect of passive ownership on the intensive learning margin. The first is the *hedging* channel: the ETF allows investors to better isolate bets on stock-specific risk-factors.[6] The second is the *market-timing* channel: the ETF

---

[5]This type of scenario, where there is a lack of volatility in both $r_{(1,2)}$ and $r_{(2,3)}$, could be the result of noise-trade demand and informed investors trading in opposite directions e.g., $\epsilon < 0$ and $z_1 > 0$, a well as a draw of v close to $\overline{v}$. In this scenario, even though prices are close to fundamentals at $t = 2$, they are still uninformative in some sense: The market maker did not learn much from the net order flow and their posterior beliefs remained close to their prior beliefs.

[6]Suppose, for example, an informed investor has a precise signal that Facebook is going to have strong earnings this quarter. When they buy Facebook stock, however, they are also exposed various systematic risk-factors. With the introduction of and rise of passive ownership this investor can go long Facebook and short the ETF to isolate their bet on firm-specific news. This may explain why hedge funds are net short almost all major ETFs (see e.g., Balchunas (2016)).

allows investors to trade directly on the systematic risk-factor. The hedging channel tends to *increase* attention to stock-specific risks, while the market-timing channel does the opposite.

In the model, passive ownership leads to less learning about systematic risk when the hedging channel dominates the market-timing channel. This happens when investors are closer to risk neutral and thus care more about trading profits than risk. An ETF lets them take risky targeted bets on volatile individual securities and they learn more about the stock-specific risk-factors. On the other hand, the ETF increases learning about systematic risk when the market timing channel dominates the hedging channel. If investors are risk averse, they care more about systematic risk because idiosyncratic risk can be diversified away. An ETF allows them trade on systematic risk directly and they want to learn even more about it.

There are also two competing forces that govern the extensive learning margin. The first is another feature of the *hedging* channel: Because passive ownership allows investors to better isolate bets on stock-specific information, it can increase the relative benefit to becoming informed. The second force is the *diversification* channel: the ETF is a less information-sensitive security (see e.g., Subrahmanyam (1991), Gorton and Pennacchi (1993)) than the individual stocks and as a result makes the uninformed investors relatively better off. For similar reasons to the intensive learning margin, the diversification channel will be stronger when risk aversion is high, while the hedging channel will be stronger when risk aversion is low.

As a result of these competing channels, the intensive and extensive margin effects of passive ownership on investors' learning are ambiguous. To resolve this ambiguity, empirical tests are needed. To this end, in the next section, I map the measures of price informativeness from Kyle (1985) and Grossman and Stiglitz (1980) to the data.

3 Mapping the model to the data

In this section, I explain my measure of passive ownership. 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. Figure 2 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.
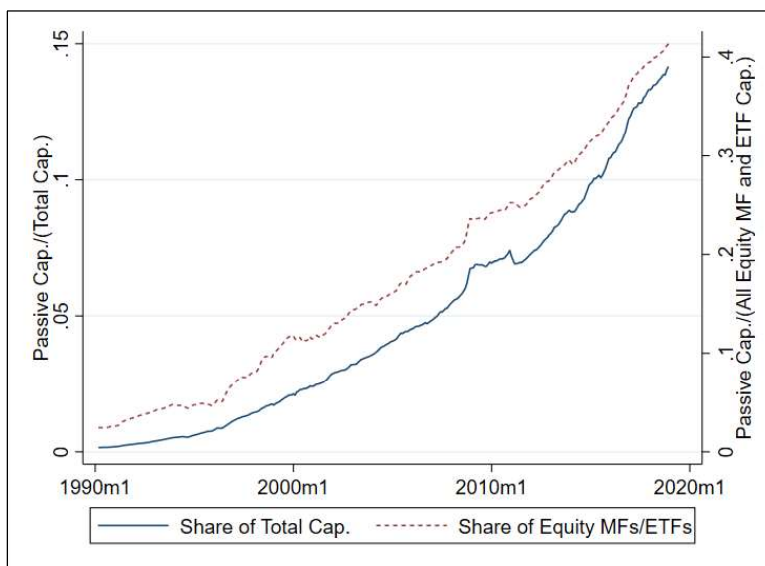

Figure 2. The rise of passive ownership: 1990-2018. Passive funds are defined as all index funds, all ETFs and all mutual funds with names that identify them as index funds. Total equity mutual fund and ETF assets is the sum of all stock holdings in the Thompson S12 data that can be matched to CRSP.

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 daily prices, returns, volume and shares outstanding 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 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.

3.3   Measure 1: Pre-earnings volume

In the models of Section 2, there are only one or two trading periods. In reality, however, there are many days between earnings announcements. Changes in trading volume, therefore, may be spread out over the month (22 trading days) before an earnings announcement. Let t denote an effective earnings announcement date. Define turnover T as total daily volume for stock i divided by shares outstanding. Then, define abnormal turnover for firm i, from event time $\tau = -22$ to $\tau = 22$ as:

$$AT_{i,t+\tau} = \frac{T_{i,t+\tau}}{T_{i,t-22}} = \frac{T_{i,t+\tau}}{\sum\limits_{k=1}^{252} T_{i,t-22-k}/252} \tag{3}$$

Where abnormal turnover, $AT_{i,t+\tau}$, is turnover divided by the historical average turnover for that stock over the past year. 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 dampening drops in trading.

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

$$AT_{i,t+\tau} = \alpha + \sum_{\tau=-21}^{22} \beta_\tau \mathbf{1}_{\{i,t+\tau\}} + e_{i,t+\tau} \tag{4}$$

The right-hand side variables of interest are a set of indicators for days relative to the earnings announcement. For example, $\mathbf{1}_{\{i,t-15\}}$ is equal to one 15 trading days before the nearest earnings announcement for stock i 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 3 plots the estimates of β_τ for $\tau = -21$ to $\tau = -2$. The estimate for $\tau = -1$ is omitted as it is about $5\times$ as large as the coefficients for $\tau = -21$ to $\tau = -2$, 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.
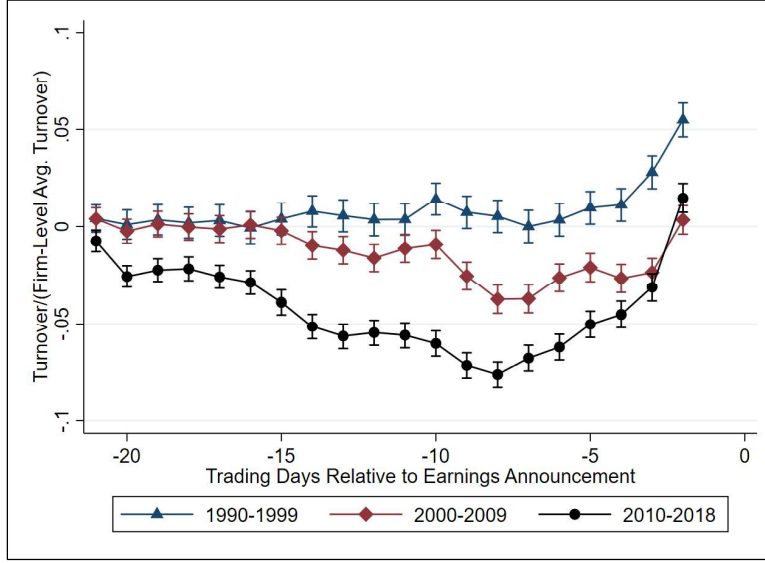
Figure 3. Decline of pre-earnings turnover. Plot of β_τ estimated from the regression:

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

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.

Figure 3 confirms that there has been a drop in trading volume throughout the month before each earnings announcement over the past 3 decades. Define cumulative abnormal

pre-earnings turnover as:

$$CAT_{i,t} = \sum_{\tau=-22}^{-1} AT_{i,t+\tau} \tag{5}$$

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 measure of price informativeness. Lower values of $CAT_{i,t}$ translate to less pre-earnings trading and in the models could be explained by a decrease in the share of informed investors and/or a decrease in the precision of informed investors' signals (Hypothesis 1). 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. The magnitude of this decrease is about 5% of CAT's whole-sample average of 22.

3.4 Measure 2: Pre-earnings drift

In the Kyle (1985)-style model, consider a stock which is going to have a high payoff at $t = 3$. As we increase the share of informed investors, or we increase the precision of informed investors' signals, the prices at $t = 1$ and $t = 2$ will be relatively higher. Empirically, this upward drift may happen over the month before the earnings announcement as informed investors aim to avoid excessive price impact.

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 before and after the earnings announcement. Figure 4 shows the average 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

18

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.
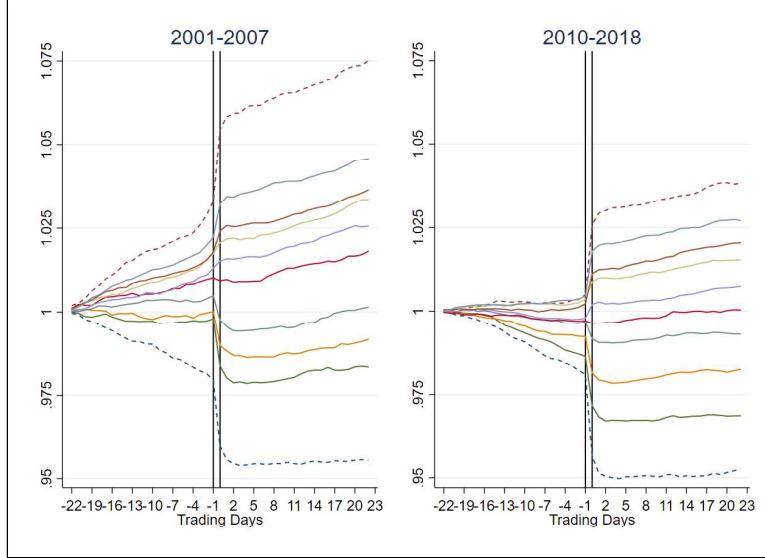

Figure 4. Decline of pre-earnings drift by SUE decile. Each quarter, I sort firms into deciles on standardized unexpected earnings. Each line represents the cross-sectional average market-adjusted return of $1 invested at t=-22. The brown dashed line represents the average for firms with the most positive 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.

Figure 4'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 models, 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 $t-1$, divided by the cumulative returns from $t-22$ to t, where t is an earnings announcement:

$$DM_{it} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t)}} & \text{if } r_t > 0 \\ \frac{1+r_{(t-22,t)}}{1+r_{(t-22,t-1)}} & \text{if } r_t < 0 \end{cases} \tag{6}$$

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

19

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.[7] I work with gross returns, rather than net returns, to avoid dividing by numbers near zero. If I had instead defined $\widehat{DM_{i,t}} = r_{t-22,t-1}/r_{t-22,t}$, the mean of $\widehat{DM_{i,t}}$ may not be well defined, as $r_{t-22,t}$ can be equal to zero.

$DM_{i,t}$ is my second main empirical measure of pre-earnings price informativeness. Lower values of $DM_{i,t}$ in the models can be explained by a lower share of informed investors and/or informed investors receiving less precise signals (Hypothesis 2). Average pre-earnings drift decreased by about -0.02 between 1990 and 2018. This drop is about 40% the size of DM's whole sample standard deviation of 0.05.

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 more volatility on earnings days, relative to all other days. In the model, when fewer investors become informed, or the precision of informed investors' signals decreases, earnings day returns become more volatile, relative to non-earnings days (Hypothesis 3). 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 t as:

$$QVS_{i,t} = 1 - \left(r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2 \right) \tag{7}$$

where r denotes a market-adjusted daily return. The numerator of the term in parenthesis is the squared earnings-day return, while the denominator is the sum of squared returns from $t-22$ to t. QVS is going to be my third main empirical measure of price informativeness. If relatively more information is being learned and incorporated into prices *on* earnings

announcement dates, we would expect smaller values of QVS.[8]

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 $QVS_{i,t}$ smaller than 0.957 imply that earnings days account for a disproportionately large share of total volatility.[9] Figure 5 plots coefficients from a regression of QVS on a set of year dummy variables for all stocks in my sample. Average QVS decreased from 92.0% in 1990 to 72.6% in 2018. This 19.4% decline is about the same size as QVS's whole-sample standard deviation of 21.0%. The Online Appendix shows that the decrease in QVS was due to a simultaneous increase earnings-day volatility and a decrease in non-earnings-day volatility.

3.6 Summary Statistics

These three measures of price informativeness are going to be the key outcome variables in all my empirical exercises. Table 1 contains details on the means, standard deviations and distributions of these measures, as well as passive ownership. Consistent with Figures 2, 3, 4 and 5, the average of all three price informativeness measures decreased between the 1990's and the 2010's, while passive ownership increased.

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. Across all three measures, passive ownership is correlated with decreased price

---

[8]This is why I define $QVS_{i,t} = 1 - \left(r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2 \right)$ instead of $\widehat{QVS_{i,t}} = r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2$. Lower values of QVS imply less informative prices, consistent with the pre-earnings volume and drift magnitude measures.

[9]All results are robust to defining QVS at the annual level i.e., defining the numerator of the term in parenthesis 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 e.g., defining $\widehat{QVS_{i,t}} = 1 - \left[\left(\sum_{j=0}^{n} r_{i,t+j}^2 \right) / \left(\sum_{\tau=-22}^{0} r_{i,t+\tau}^2 \right) \right]$ for $n = 1, \dots, 5$.

Figure 5. 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: $QVS_{i,t} = 1 - \left(r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2 \right)$, 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.

informativeness.

4.1 Pre-earnings turnover

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

$$CAT_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{8}$$

where cumulative abnormal pre-earnings turnover, $CAT_{i,t}$, is defined in Equation 5. Controls in $X_{i,t}$ include time since listing (which I call firm age), one-month lagged market capitalization, returns from t-12 to t-2 (the returns typically used to form momentum portfolios), 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

		25%	50%	Mean	75%	St. Dev.
CAT		13.33	19.02	22.57	26.85	14.87
DM		0.956	0.979	0.966	0.992	0.042
QVS	1990-1999	0.896	0.969	0.913	0.994	0.136
Passive		0.002	0.005	0.008	0.012	0.011
CAT		14.53	18.76	21.23	24.12	12.40
DM		0.936	0.969	0.952	0.988	0.054
QVS	2010-2018	0.605	0.879	0.764	0.978	0.263
Passive		0.014	0.067	0.077	0.120	0.070
CAT		13.84	19.03	22.15	25.87	14.30
DM		0.942	0.973	0.956	0.989	0.053
QVS	All Years	0.791	0.944	0.845	0.990	0.215
Passive		0.004	0.016	0.037	0.050	0.052

Table 1 Summary Statistics. Cross-sectional means, standard deviations and distributions of price informativeness and passive ownership.

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)).[10]

Equation 8 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 2 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

---

[10]Time since listing, market capitalization and past returns are computed using CRSP. 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. CAPM beta and R-squared are from the WRDS beta suite.

of restricting only to 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. Going forward, I refer to this specification, with equal weights, all the firm-level controls and fixed effects as the baseline specification.

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.[11] 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 market capitalization at the end of the previous quarter. Using value weights, instead 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 2 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.

---

[11]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.

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:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \qquad (9)$$

where $DM_{i,t}$ is defined as in Equation 6 and all the controls and fixed effects are the same as in Equation 8. The regression results are in Table 3. 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 instead of equal 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 3 Passive ownership and pre-earnings drift. 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:

$$QVS_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{10}$$

where QVS is defined in Equation 7 and all the controls and fixed effects are the same as in Equation 8. The regression results are in Table 4. The baseline specification (column 3) implies that a 15% increase in passive ownership would lead to a decrease in QVS of 6.1%. 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.

	(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 4 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.

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

As discussed in the introduction, passive ownership is correlated with decreased information production about the underlying firms. This raises the concern of reverse causality: maybe passive ownership increased the most in stocks that (for other reasons) had the largest decrease in price informativeness. To confirm that passive ownership is the cause of the decrease in price informativeness, 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 mutual funds and ETFs buy the stock.

I obtain daily S&P 500 index constituents from Compustat between 1990 and 2017. 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. Because the S&P 500 is comprised of large firms the additions are almost exclusively in the fourth and fifth market capitalization quintiles of their 3-digit SIC industry.

I also form a second control group of firms meeting the same selection criteria (3-digit SIC industry and industry market capitalization quintile), but that 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 in 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:

$$Passive_{i,t} = \alpha + \beta_1 Post_{i,t} + \beta_2 Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \tag{11}$$

Here, $Treated_{i,t}$ is equal to one if the firm was added to the S&P 500 and zero otherwise.[12] The level of observation is firm-quarter-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. There are fixed effects for each firm-cohort group to account for the fact that a firm-quarter observation can appear in multiple cohorts. There are also fixed effects for the month of index rebalancing to account for time-series trends. $Treated_{i,t}$ is not included in Equation 11 because each firm can only be in one of the three treatment categories within each cohort group, so this term would be washed out by the fixed effects.

Figure 6 shows the level of passive ownership for the control firms and treated firms around the month of index addition. Within each cohort group, I subtract the average *level* of passive ownership 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.

S&P 500 index additions do not always coincide with the end of a calendar quarter. Given that the S12 data I use to quantify passive ownership is quarterly, I do not know the level of passive ownership exactly 3 months before, in the month of and 3 months after index addition for all treated and control firms. In constructing Figure 6, I fix the level of passive ownership at its last reported level each month between quarter-ends. Although it appears as though passive ownership increases slowly around index addition, this is partly a function of averaging across observations with differences in time until the first set of post-index-addition S12 filings.

---

[12]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.
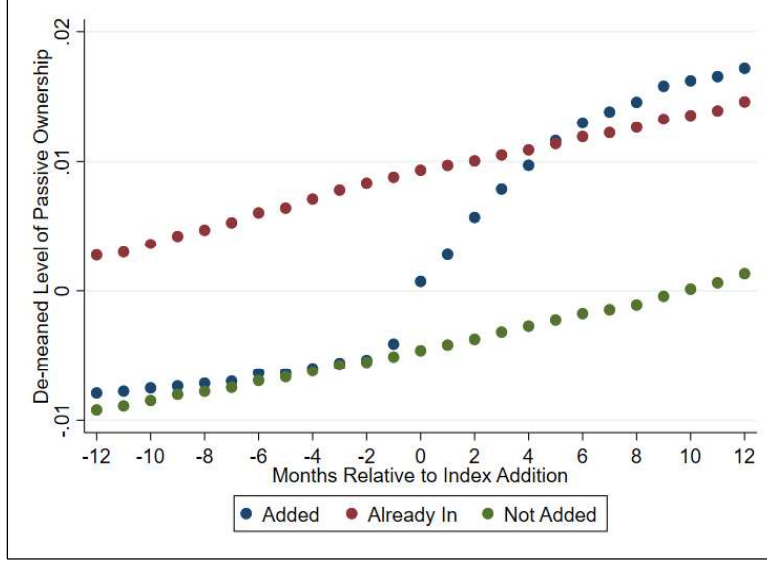
Figure 6. S&P 500 index addition and changes in passive ownership. Average level of passive ownership for control firms out of the index ("Not Added"), control firms in the index ("Already In") and treated firms ("Added"). Passive ownership is demeaned within each cohort group i.e., within each group 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:

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

$$Outcome_{i,t} = \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t} \tag{12}$$

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

Where $Outcome_{i,t}$ is pre-earnings turnover, drift, or earnings days' share of volatility. The fixed effects are the same as in Equation 11. 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 the pre-addition increase in passive ownership visible in Figure 6. This increase may occur because S&P announces index addition before it actually occurs e.g., for Tesla index addition was announced on 11/16/2020 but it was not added until 12/21/2020.

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:

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

$$Outcome_{i,t} = \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t} \qquad (13)$$

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

Here, Passive 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 the treated firm is added to the index. If at the time of index addition there are not matched control firms both in and out of the index, I use the average Passive Gap$_{i,t}$ for all firms that year. Passive 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%. The uninteracted term Passive Gap$_{i,t}$ is not included in Equation 13 because it is constant within each cohort group and would be washed out by the fixed effects. Following Coles et al. (2020), I double cluster standard errors at the firm and quarter level in Equations 12 and 13.

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 5 contains the regression results. Panel A examines the effect of index addition on price informativeness using the binary treatment specification. Column 1 is the first stage regression.[13] The associated F-statistic is very large, which is not surprising given the increase in passive ownership pictured in Figure 6.[14] The coefficient on $Post \times Treated$

---

[13]Column 1 only uses observations with non-missing pre-earnings abnormal turnover data. The number of observations and thus the first-stage regression is slightly different for each of the three measures of price informativeness. The results, however, are quantitatively similar in each case, so I only report this first stage regression to avoid redundancy.

[14]Because I am using both $Post$ and $Post \times Treated$ as instruments for passive ownership, the time trend *and* the treatment effect in Figure 6 are driving the large magnitude of the F-statistic in Table 5. In a regression of passive ownership on $Post$, $Post \times Treated$ and the fixed effects, both terms are individually statistically significant, with $Post$ having a t-Statistic of 18 and $Post \times Treated$ having a t-Statistic of 15.

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* × *Treated* as instruments for passive ownership. The effect is negative, consistent with the cross-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 (RF) 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 4-5 and 6-7 repeat columns 2-3 for the pre-earnings drift and QVS. For the drift, the IV estimate is the same sign as the cross-sectional coefficient of -0.05, but is about three times as large. This is also true for QVS, with the IV estimate approximately tripling the baseline estimate of -0.41. 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.

The reduced-form regressions for the drift and QVS are the same sign as the baseline estimates, but statistically insignificant. 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 associated with index addition 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 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 i.e., Passive Gap$_{i,t}$ should be informative about anything other than the *sign* of the treatment effect.

Panel B repeats panel A, but using the gap in passive ownership between the two sets of matched control firms interacted with the treatment dummy, along with $Post_{i,t}$, as instruments for passive ownership. As with the binary treatment specification, the first stage is economically large and statistically significant. The IV estimates in Panel B are quantitatively similar to the IV estimates Panel A. With the continuous instrument, however, it more straightforward to compare the magnitude of the reduced-form estimates with the baseline cross-sectional results. The estimate for the decline in pre-earnings volume is about twice as large as the baseline of -11.2, while the estimates for the pre-earnings drift and QVS are

each about half as large as the baseline estimates of -0.05 and -0.40.

| | First Stage | Panel A: Binary Instrument | | | | | |
| | | Pre-Earnings Volume | | Pre-Earnings Drift | | QVS | |
	First Stage	IV	RF	IV	RF	IV	RF
Post x Treated	0.0135***		-0.621**		-0.00194		-0.00591
	(0.001)		(0.253)		(0.001)		(0.005)
Passive Ownership		-15.41		-0.172***		-1.513***	
		(10.270)		(0.052)		(0.163)	
Observations	188,606	188,606	188,606	190,940	190,940	192,469	192,469
F-statistic	258						

| | First Stage | Panel B: Continuous Instrument | | | | | |
| | | Pre-Earnings Turnover | | Pre-Earnings Drift | | QVS | |
	First Stage	IV	RF	IV	RF	IV	RF
Post x Treated x Passive Gap	0.550***		-24.90***		-0.028		-0.195
	(0.045)		(8.544)		(0.042)		(0.164)
Passive Ownership		-15.57		-0.168***		-1.502***	
		(10.220)		(0.052)		(0.162)	
Observations	188,606	188,606	188,606	190,940	190,940	192,469	192,469
F-statistic	385						

Table 5 Effects of S&P 500 index addition on price informativeness. Estimates from:

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

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

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

Column 1 in each panel is a first-stage regression. Columns 2, 4 and 6 are instrumental variables regressions. Columns 3, 5 and 7 are reduced-form regressions. Panel A contains regressions from the binary treatment specification, while Panel B contains regressions from the continuous treatment specification. FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are 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.

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 market capitalization of 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.[15] 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.[16] The treated firms are those that were over the threshold and ended up switching, while the control firms are those that were above the threshold and stayed in the 1000. 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%, so including these years

---

[15]I would like to thank the authors for sharing their replication code with me. 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), Großner (2018), Ben-David et al. (2019) and Heath et al. (2021).

[16]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.

leads to a weaker first stage. One reason for this is that the two largest Russell 1000/2000 funds, IWB and IWM, were not launched until May 2000. (2) Under the bandwidth regime, switching is harder to predict/manipulate, so switching is less likely to be front-run by other institutional investors.

Figure 7 compares the level of passive ownership around the index rebalancing date between the treated and control group. Within each cohort, I subtract the mean level of passive ownership. The pre-addition changes and levels of passive ownership are similar between both groups. Unlike S&P 500 index additions, Russell reconstitutions always coincide exactly with the end of a calendar quarter. Because of this, Figure 7 only plots data points for months with S12 filings i.e., the last month of each calendar quarter.
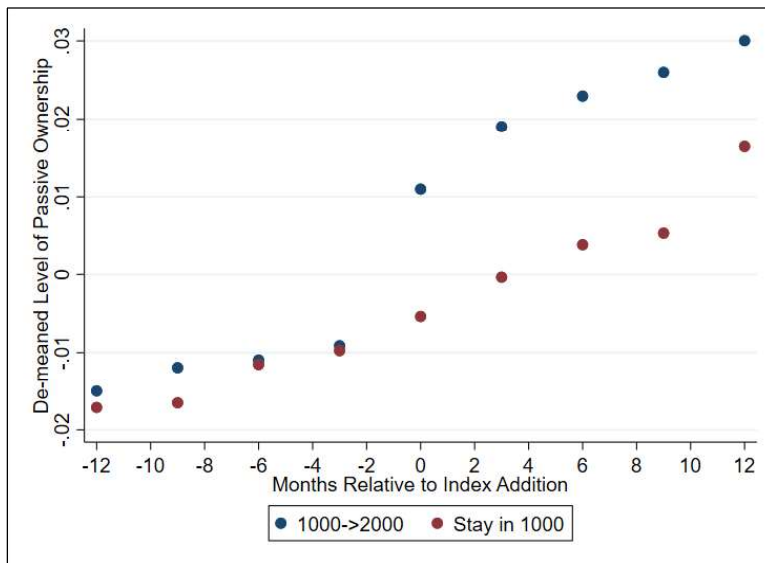
Figure 7. Russell 1000/2000 reconstitution and changes in passive ownership. Average level of passive ownership for firms that stay in the Russell 1000 (control firms) and firms that moved from the Russell 1000 to the Russell 2000 (treated firms). Passive ownership is demeaned within each cohort.

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 of the 1000th ranked firm in March, before index rebalancing (2) the fixed effects are identical, but cohort group is now defined only by month of index rebalancing and (3) the time period is different, as I use Russell reconstitutions between 2007 and 2018.

Table 6 contains the regression results. The first-stage results are positive, with a large F-statistic. The estimated coefficient of 1.42%, however, understates the *total* change in passive ownership. In Figure 7, 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 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 a decrease in QVS. The IV estimates of -22.89 and -0.12 for the volume and drift specifications are significantly larger than the estimates from Tables 2 and 3, while the IV estimate of -0.22 for QVS is about half the size of those in Table 4.

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   Mechanisms

Motivated by the richer model in the Online Appendix, in this Section I provide evidence on a learning-based mechanism through which passive ownership could decrease price informativeness. In the cross section, passive ownership is correlated both with decreased information production by sell-side analysts and decreased downloads of SEC filings. I then proceed to rule out several alternative explanations by showing that (1) the trends and cross-sectional regression results are specific to earnings announcement days (2) the results are robust to using an ex-ante measure of earnings uncertainty and (3) the results are not driven by changes in the nature of information releases, but rather changes in the market's response to information of a given size, especially for firms with higher levels of passive ownership.

| | | Panel A: Binary Instrument | | | | | |
| | First Stage | Pre-Earnings Turnover | | Pre-Earnings Drift | | QVS | |
		IV	RF	IV	RF	IV	RF
Post x Treated	0.0142***		-0.742		-0.00615		-0.000894
	(0.005)		(0.502)		(0.004)		(0.015)
Passive Ownership		-22.89***		-0.116***		-0.216	
		(8.133)		(0.040)		(0.160)	
Observations	5,893	5,893	5,893	6,174	6,174	6,175	6,175
F-statistic	145						

| | | Panel B: Continuous Instrument | | | | | |
| | First Stage | Pre-Earnings Turnover | | Pre-Earnings Drift | | QVS | |
		IV	RF	IV	RF	IV	RF
Post x Treated	0.759***		-22.37		-0.116		-0.224
x Passive Gap	(0.200)		(20.560)		(0.140)		(0.640)
Passive Ownership		-22.60***		-0.111***		-0.221	
		(7.803)		(0.038)		(0.161)	
Observations	5,893	5,893	5,893	6,174	6,174	6,175	6,175
F-statistic	158						

Table 6 Effects of Russell 1000/2000 index reconstitution on price informativeness. Estimates from:

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

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

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

Column 1 in each panel is a first-stage regression. Columns 2, 4 and 6 are instrumental variables regressions. Columns 3, 5 and 7 are reduced-form regressions. Panel A contains regressions from the binary treatment specification, while Panel B contains regressions from the continuous treatment specification. FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

6.1 Information gathering

6.1.1 Information production by sell-side analysts

Passive managers, as well as investors in passive funds, lack strong incentives to gather and consume firm-specific information. Passive funds trade on mechanical rules, such as S&P 500 index membership (SPY), or the 100 lowest volatility stocks in the S&P 500 (SPLV). Given that these trading strategies are implemented on public signals and are well diversified, they do not necessarily require accurate private forecasts of firm fundamentals. Sell-side analysts may respond to these incentives by producing less or lower quality information

about stocks with higher levels of passive ownership.

On the other hand, as a stock becomes more mispriced, the return to gathering fundamental information increases. Given that passive ownership reduces price informativeness, analysts might decide to produce more information about the underlying stocks. Ex-ante, it is not obvious which of these effects will dominate in equilibrium, so to distinguish between these alternative stories, I run the following regression:

$$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{14}$$

where $Outcome_{i,t}$ will be measures of information production by sell-side analysts. The sample is all quarterly earnings announcements in IBES, further restricting to observations that can be (1) matched to CRSP (2) have at least 3 estimates of earnings-per-share (3) have a non-missing value for realized earnings per share and (4) have a non-missing closing price on the last trading day before the earnings announcement in CRSP. Within each forecast period, I take the last statistical period i.e., the last set of estimates before the earnings information is released.

Table 7 contains the results. Column 1 shows that higher passive ownership is correlated with lower analyst coverage. This is consistent with Israeli et al. (2017) and Coles et al. (2020), who show that ETF ownership is negatively correlated with the numbest of analyst estimates. Column 2 shows that passive ownership is correlated with a larger standard deviation of analyst estimates. This increased forecast dispersion is evidence of more uncertainty about the fundamental value of these firms (see e.g., Diether et al. (2002), Zhang (2006)).

One concern, however, is the increased standard deviation of forecasts is a mechanical function of the decrease in coverage documented in Column 1. To address this, I construct a measure of analyst (in)accuracy which accounts for the increase in dispersion: the absolute difference between realized earnings and the mean estimate of earnings, divided by the standard deviation of analysts' estimates. If analysts are producing lower quality information, we would expect their forecasts to be less accurate, even when accounting for the average increase in uncertainty.

Column 3 shows that analysts' forecasts are less accurate for firms with more passive ownership. As a robustness check, Column 4 uses an alternative measure of inaccuracy: the absolute difference between realized earnings and the mean estimate of earnings, divided by

the closing price on the last trading day before the earnings information was made public. This alternative inaccuracy measure is also positively correlated with passive ownership. Normalizing earnings estimate inaccuracy by the price is not my preferred specification, however, as average price informativeness declined over the sample and price informativeness is negatively correlated with passive ownership.

Columns 5 and 6 restrict to the subset of announcements which are covered by analysts who update their forecasts at least once between when they initiate coverage for a fiscal period and when earnings information is released. Columns 5 shows that analysts update their estimate of earnings less frequently for stocks with more passive ownership. In a similar vein, Column 6 shows that the average time between updates is higher for stocks with high passive ownership.

	Num. Est. (1)	SD(Est.) (2)	Dist./SD(Est.) (3)	Dist./P (4)	Updates (5)	Time (6)
Passive Ownership	-8.451***	0.681***	1.843***	0.158**	-0.326***	0.269**
	(1.344)	(0.171)	(0.419)	(0.075)	(0.095)	(0.105)
Observations	214,304	214,304	214,304	214,304	132,195	132,195
R-Squared	0.79	0.641	0.125	0.257	0.255	0.544
Mean	8.804	0.0915	2.244	0.026	2.24	3.765
St. Dev.	6.054	0.41	2.984	0.261	0.454	0.853

Table 7 Passive ownership and information production. Estimates of β from:
$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t}$
Num. Est. is the number of analyst estimates, SD(Est.) is the standard deviation of analyst estimates, Dist. is the absolute distance between realized earnings per share and the mean estimate of earnings per share, P is the last pre-earnings announcement price, Updates is the average number of analyst updates within each forecasting period and Time is the average number of days between analyst updates within each forecasting period. 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. All columns contain year-quarter fixed effects, ϕ_t, quarter-of-year fixed effects, ζ_q and firm fixed effects ψ_i. Standard errors double clustered at the firm and year-quarter level in parenthesis. The last two rows of the table present the mean and standard deviation of the left-hand-side variables.

The findings in Table 7 seem at odds with results in the previous section: it is well known, for example, that when firms are added to the S&P 500 they receive *increased* analyst coverage. Decreased incentives to gather or produce firm-specific information could

still, however, explain those results. For example, suppose analysts know that after a firm is added to the S&P 500 index, a larger share of its investors are holding it as a part of a well-diversified portfolio. They may, therefore, choose not to expend the effort required to produce an equally accurate measure of firm fundamentals as they would if their clients were taking isolated bets on the stock.

6.1.2 Downloads of SEC filings

Sell side analysts are not necessarily investors themselves. A more direct test of investor attention is to examine the downloads of SEC filings (see e.g., Loughran and McDonald (2017)). More downloads implies increased gathering of fundamental information. If fewer investors are paying attention to stocks with more passive ownership, it should be negatively correlated with downloads of SEC filings. The timing of downloads relative to earnings announcements, however, is not obvious. Attentive investors may download the reports (1) right before earnings are released to e.g., make a bet ahead of the announcement (2) on the earnings announcement date to e.g., bet on the announcement news or (3) some time after earnings are released to e.g., bet on a re-interpretation the announcement news.

Rather than trying to distinguish between these stories, I perform a more general test. At the stock/month level, I ask whether stocks with more passive ownership have fewer downloads of SEC filings than stocks with less passive ownership. To this end, I run the following regression:

$$Downloads_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t} \tag{15}$$

where $Downloads_{i,t}$ are the number of non-robot downloads, measured using the method in Loughran and McDonald (2017) and obtained from their website. I match the downloads data to the CRSP/Compustat merged database on CIK. The sample runs from 2003-2015, excluding the data lost/damaged by the SEC from 9/2005-5/2006. Downloads have been trending up over time and there is a large increase in downloads – relative to the unconditional firm-level average – on earnings announcement dates.

Table 8 contains the results. In Column 1, which only has firm/time fixed effects, the relationship is positive and statistically significant. This could be explained, however, by the fact that high passive firms are larger on average and large firms attract more investor

attention. Column 3 adds in all the firm level controls, including one-month lagged market capitalization and total institutional ownership. This flips the sign, making the coefficient negative and statistically significant, evidence that passive ownership is correlated with less investor attention. This result is consistent with Coles et al. (2020), who show that ETF ownership is negatively correlated with downloads of SEC filings.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	0.526***	-0.00485	-0.623***	-0.756	-1.360**
	(0.164)	(0.171)	(0.185)	(0.568)	(0.625)
Observations	608,722	518,089	518,089	518,089	518,089
R-Squared	0.788	0.806	0.808	0.889	0.893
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table 8 Passive ownership and downloads of SEC filings. Estimates of β from:
$Downloads_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t}$
$Downloads$ is the number of non-robot downloads from Loughran and McDonald (2017). 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. All columns contain year-quarter fixed effects, ϕ_t, quarter-of-year fixed effects, ζ_q and firm fixed effects ψ_i. Standard errors double clustered at the firm and year-quarter level in parenthesis.

6.2 Ruling out alternative explanations

6.2.1 Placebo tests on stylized facts

One concern is that Section 3's 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 QVS for 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 or QVS for the

placebo earnings dates. I repeat this exercise using randomly selected dates as placebo earnings announcements and also 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. Like earnings announcements, these are days where large quantities of information are released, but it is not specific to any particular firm. 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 released around earnings announcements.

6.2.2 Placebo tests on cross-sectional regression results

A related concern is that cross-sectional regression results of Section 4 are not specific to earnings announcement days. An alternative explanation is that passive ownership increases overall volatility (see e.g., Ben-David et al. (2018)) and given the upward trend in passive ownership, this could create a spurious correlation between passive ownership and QVS. Recall $QVS_{i,t} = 1 - \left(r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2 \right)$, so $\sum_{\tau=-22}^{0} r_{i,t+\tau}^2$ contains volatility relatively further in the past than $r_{i,t}^2$. If increases in passive ownership are persistent at the firm level, QVS might mechanically decrease 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, in the Online Appendix, I perform three additional placebo tests.

As before, 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 negative relationship between passive ownership and QVS, but the magnitude is 1/20th as large as the coefficient in Table 4. 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 releases.

6.2.3 Regime shifts

Two additional threats to identification are (1) Regulation Fair Disclosure (Reg FD), passed in August 2000, which reduced the 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).[17]

6.2.4 Option implied volatility

Because DM and QVS are computed using earnings-day returns, they are *ex-post* measures of uncertainty. If fewer investors are becoming informed or investors' signals have become less precise, however, we would also expect an increase in *ex-ante* uncertainty. One way to measure ex-ante uncertainty about fundamentals is via the cost of options exposed to earnings announcement risk (see e.g., Dubinsky et al. (2006)). Following Kelly et al. (2016), I compute a version of their *Implied Volatility Difference* (IVD) measure to quantify how much more expensive options that span earnings announcements are, relative to options that expire the month before/after the announcement.[18]

Letting τ denote an earnings announcement date, I identify regular monthly expiration dates a, b and c, such that $a < \tau < b < c$. The final variable of interest, the implied volatility difference, is defined as:

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right) \tag{16}$$

higher values of $IVD_{i,\tau}$ imply that options which span earnings announcements are relatively more expensive i.e., there is more ex-ante uncertainty about the earnings news.[19] The Online Appendix shows that average IVD is positive and has increased by about 0.05 over the past

---

[17]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.

[18]Although I broadly follow the method in Kelly et al. (2016) to construct IVD, the Online Appendix contains step-by-step details on my exact variable construction procedure.

[19]One concern with this definition of IVD is that subtracting the average of $\overline{IV}_{i,a}$ and $\overline{IV}_{i,c}$ from $\overline{IV}_{i,b}$ accounts for firm-specific time trends in implied volatility, but not *level* differences in implied volatility across firms. All the results that follow are qualitatively unchanged using $I\tilde{V}D_{i,\tau} = \overline{IV}_{i,b}/\frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right)$.

25 years. This is evidence that there is more uncertainty about fundamentals before earnings announcements now than there was in the late 1990's.

The natural next step is to determine the relationship between IVD and passive ownership. If passive ownership is leading to less information gathering, we would expect it to be correlated with more pre-earnings announcement uncertainty. Table 9 contains the results. Consistent with this hypothesis, IVD is positively correlated with passive ownership. In terms of magnitudes, a 15% increase in passive ownership implies about a 0.02 higher IVD on average. This is an economically large effect, at about 40% of the increase in average IVD over the whole sample.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	0.129***	0.141***	0.115***	0.166***	0.169***
	(0.028)	(0.029)	(0.030)	(0.039)	(0.035)
Observations	115,213	108,823	108,823	108,823	108,823
R-Squared	0.279	0.285	0.291	0.418	0.425
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table 9 Passive ownership and IVD for earnings announcements. Estimates of β from:
$IVD_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + e_{i,t}$
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. All columns contain year-quarter fixed effects, ϕ_t, quarter-of-year fixed effects, ζ_q and firm fixed effects ψ_i. Standard errors double clustered at the firm and year-quarter level in parenthesis.

6.2.5 Earnings responses

As shown in Section 3, there has been a trend toward increased volatility on earnings-announcement days. A possible explanation is that the nature of news releases has changed over time. In addition, Section 4 shows that firms with more passive ownership have relatively more volatility around earnings announcements. Again, a possible alternative explanation is that stocks with more passive ownership have meaningfully different types of news than stocks with low passive ownership.

43

To rule out these alternatives, I use an earnings-response regression based on Kothari and Sloan (1992) to quantify the market's reaction to a standardized measure of earnings news. In the Online Appendix, I show earnings responses increased by a factor of about $3\times$ over the past 30 years i.e., the market reaction to earnings news of a *given size* is about 3 times as large now as it was in the early 1990's. This trend is one explanation for the time-series decrease in average DM and QVS.

To understand the relationship between passive ownership and earnings responses, I run the following regression:

$$
\begin{aligned}
r_{i,t} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 \left(SUE_{i,t} \times Passive_{i,t} \right) \\
+ \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + \epsilon_{i,t}
\end{aligned}
\tag{17}
$$

where $r_{i,t}$ denotes the market-adjusted return on the effective quarterly earnings date i.e., the first day investors could trade on earnings information. Results in this subsection are similar when instead using cumulative returns in windows of up to 5 days after the earnings-announcement. $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t}-E_{i,t-4})}$, where $E_{i,t}$ is earnings-per-share from the IBES unadjusted detail file. The numerator is the year-over-year (YOY) earnings growth, while the denominator is the standard deviation of YOY earnings growth over the past 8 quarters.[20] I also run versions of Equation 17: (1) breaking SUE into positive and negative components and (2) decomposing the earnings news into a systematic and idiosyncratic component using the method in Glosten et al. (2021).[21]

In a way, Equation 17 is a conditional version of Equations 9 and 10 – the cross-sectional DM and QVS regressions – where I am conditioning on the size of earnings news. If passive ownership itself is driving my results, rather than differences in the nature of earnings news between high and low passive firms, we would expect γ in Equation 17 to be positive.

Table 10 contains the regression results. Columns 1-3 are a sanity check and do not include the interaction terms with passive ownership. Everything is consistent with common-

---

[20] I compute SUE this way, following Novy-Marx (2015), because it avoids (1) using prices as an input, whose average informativeness has changed over time and (2) using analyst estimates of earnings as an input, whose average accuracy has also changed over time. As a result, the average absolute value of $SUE_{i,t}$ is roughly constant over my sample, except for large spikes during the tech boom/bust as well as during the global financial crisis.

[21] This is done by regressing SUE on market-wide SUE and SIC-2 industry-wide SUE in five year rolling windows. The systematic component of earnings is the predicted value from this regression, while the idiosyncratic component is the residual.

sense intuition: (1) SUE is positively correlated with earnings-day returns (2) this is true individually both for the positive and negative components of SUE and (3) this is also true individually for the positive/negative components of SUE when decomposing earnings news into systematic and idiosyncratic components.

Columns 4-6 add the level of passive ownership, as well as all the interaction terms with passive ownership to Columns 1-3. In Column 4, consistent with passive ownership driving the fall of DM and QVS, γ is positive and economically large. The estimates in Column 5 imply that firms with a high share of passive ownership are more responsive to earnings news, especially if that news is negative. Column 6 shows that firms with higher levels of passive ownership are especially responsive to negative idiosyncratic news. This increased responsiveness to earnings news is one explanation for why DM and QVS are relatively lower for firms with more passive ownership.

7 Conclusion

The first goal of this paper is to define measures of price informativeness which are easy to bring to the data. I leverage two canonical models of trade under asymmetric information to motivate three new measures of price informativeness based on trading volume, returns and volatility around earnings announcement dates. 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. These results confirm that passive ownership is decreasing price informativeness and not vice-versa.

Finally, I provide evidence on a mechanism for the empirical results related to investors' learning behavior: Stocks with more passive ownership have both a lower quantity and quality of information production by sell-side analysts. This decrease in attention is not limited to sell-side analysts, as passive ownership is also correlated with decreased downloads of SEC filings.

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)
SUE	0.407***			0.359***		
	(0.0229)			(0.0247)		
SUE>0		0.653***			0.668***	
		(0.0378)			(0.0376)	
| SUE<0 |		-0.297***			-0.228***	
		(0.0203)			(0.0221)	
Sys. SUE>0			0.372*			0.600***
			(0.1940)			(0.1790)
|Sys. SUE<0|			-0.191*			-0.218
			(0.1060)			(0.1380)
Idio. SUE>0			0.658***			0.660***
			(0.0358)			(0.0366)
|Idio. SUE<0|			-0.289***			-0.219***
			(0.0240)			(0.0240)
SUE x Passive				0.917*		
				(0.4660)		
SUE>0 x Passive					-0.284	
					(0.8250)	
| SUE<0 x Passive | x Passive					-1.385***	
					(0.3730)	
Sys. SUE>0 x Passive						-2.835
						(2.3700)
| Sys. SUE<0 | x Passive						0.702
						(1.8820)
Idio. SUE>0 x Passive						-0.0738
						(0.7760)
| Idio. SUE<0 | x Passive						-1.429***
						(0.4150)
Observations	332,558	332,558	332,558	332,558	332,558	332,558
R-squared	0.062	0.063	0.063	0.062	0.063	0.063

Table 10 Passive ownership and earnings responses. Estimates from:

$$r_{i,t} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 \left(SUE_{i,t} \times Passive_{i,t}\right)$$
$$+ \gamma X_{i,t} + \phi_t + \psi_i + \zeta_q + \epsilon_{i,t}$$

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. All columns contain year-quarter fixed effects, ϕ_t, quarter-of-year fixed effects, ζ_q and firm fixed effects ψ_i. Standard errors double clustered at the firm and year-quarter level 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).

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*.

Diether, K. B., Malloy, C. J., and Scherbina, A. (2002). Differences of opinion and the cross section of stock returns. *The Journal of Finance*, 57(5):2113–2141.

Dubinsky, A., Johannes, M., Carr, P., Chen, J., Chernov, M., Duffie, D., Glasserman, P., Heidari, M., Hull, B., Jagannatahn, R., Jones, C., Knight, J., and Linnainmaa, J. (2006). Fundamental uncertainty , earning announcements and equity options. *Wall Street Journal*, (August 2015).

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

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*.

Israeli, D., Lee, C. M., and Sridharan, S. A. (2017). Is there a dark side to exchange traded funds? An information perspective. *Review of Accounting Studies*, 22(3).

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

Kelly, B., Pástor, u., and Veronesi, P. (2016). The Price of Political Uncertainty: Theory and Evidence from the Option Market. *Journal of Finance*, 71(5).

Kothari, S. P. and Sloan, R. G. (1992). Information in prices about future earnings. Implications for earnings response coefficients. *Journal of Accounting and Economics*, 15(2-3).

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

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

Loughran, T. and McDonald, B. (2017). The use of edgar filings by investors. *Journal of Behavioral Finance*, 18(2):231–248.

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).

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). Earnings management. *Accounting horizons*, 3(4):91.

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).

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).

Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, 61(1):105–137.