

Passive Ownership and Price Informativeness

MARCO SAMMON *

December, 2022

ABSTRACT

I show that passive ownership negatively affects the degree to which stock prices anticipate earnings announcements. Estimates across several research designs imply that the rise in passive ownership over the last 30 years has caused the amount of information incorporated into prices ahead of earnings announcements to decline by roughly 16%. This effect occurs in part because passive owners collect less firm-specific information ahead of earnings announcements and limits to arbitrage prevent non-passive investors from fully offsetting this behavior.

Keywords: Passive ownership, Price informativeness.

JEL classification: G12, G14.

*Harvard Business School. I would like to thank my dissertation committee: Scott Baker, Alex Chincio, Ravi Jagannathan, Robert Korajczyk and Dimitris Papanikolaou. I would also like to thank Jennie Bai (discussant), Nick Bloom, Caitlin Dannhauser (discussant), Ian Dew-Becker, Simon Gloßner, Robin Greenwood, Sam Hanson, Mina Lee (discussant), Matthew Pritsker, Adi Sunderam, Emil Siriwardane, Martin Thyrsgaard, and Luis Viceira for helpful comments.

1 Introduction

Passive investing through index mutual funds and ETFs plays an increasingly large role in U.S. capital markets. From 1990 to 2018, the share of U.S. equities held by passive investors rose from less than 1% to almost 15%. There is still considerable debate about the costs and benefits of passive investment vehicles. Proponents of these instruments argue that they provide investors with access to a range of diversified portfolios at low costs due to a combination of lower fees, decreased turnover and greater tax efficiency (Wurgler (2010), Madhavan (2014), Madhavan (2016)). On the other hand, the growth of passive investing has raised concerns that capital market prices have become less informative, thereby distorting capital allocations (Brogaard et al., 2019). The general argument for this view is that passive investors pay less or no attention to the underlying securities and therefore their prices do not reflect all available information. Research on the relationship between passive ownership and price informativeness has drawn mixed conclusions, in part because information (i.e., the true fundamental value of the stock) is hard to measure.

In this paper, I bring new evidence to bear on this debate by studying how passive ownership affects the incorporation of information into prices in narrow windows around earnings announcements. My approach is motivated by early studies of market efficiency (Ball and Brown, 1968) documenting that a substantial portion of earnings news is incorporated into prices prior to the actual announcement. The usual interpretation of these findings is that private information is collected by market participants ahead of earnings announcements. I use this logic to test the following hypothesis: do stocks with more passive ownership have less of their earnings information incorporated into their prices ahead of earnings announcements?

I measure the amount of information incorporated into prices prior to announcements in two ways. In both cases, I examine prices in the month leading up to and including a firm's earnings announcement date. First, I compute the fraction of the total gross return (the pre-earnings drift magnitude, DM) that occurs prior to the earnings release.¹ Second, I compute the fraction of total volatility (the quadratic variation share, QVS) that occurs prior to the earnings release. I interpret lower values of both as an indication that less private

¹Weller (2018) uses a similar measure to quantify whether algorithmic trading activity deters information acquisition ahead of earnings announcements.

information was collected ahead of the earnings announcement.

Leveraging these measures, I establish several new facts about passive ownership and price informativeness before earnings announcements using the cross-section of U.S. equities from 1990 to 2018. My first main finding is that average price informativeness declined steadily over the past 30 years, mirroring the aggregate rise of passive ownership. In 1990, 92.1% of return volatility in the month leading up to and including an earnings announcement occurred prior to the release. By 2018, this number had declined to 72.4%. The average share of the total gross return occurring over the same window (DM) displays a similar downward trend. Both measures indicate that on average, less information is being incorporated into prices ahead of earnings announcements.

These aggregate patterns are mirrored in the cross-section of U.S. stocks. Through a series of panel regressions, I establish a robust negative relationship between pre-earnings announcement price informativeness and the fraction of individual firms' shares outstanding held by passive investors. My preferred regression estimates imply that a stock in the 90th percentile of passive ownership in 2018 has 9.1 pp less of its return volatility occur ahead of earnings announcements relative to a stock in the 10th percentile of passive ownership in 2018. For reference, the difference in passive ownership share between these two percentiles is 23%, slightly larger than the value-weighted average increase in passive ownership over my entire sample. Once again, the results for DM corroborate these findings. As additional supporting evidence for these results, I show that options markets internalize the relationship between passive ownership and decreased pre-earnings announcement price informativeness.

These reduced form cross-sectional correlations do not, however, conclusively establish a causal link between passive ownership and pre-earnings announcement price informativeness. An alternative interpretation is that causality runs the other way. For instance, passive vehicles may be more likely to own firms with larger market capitalizations. Larger firms are also more complex, so perhaps less information is incorporated into their prices ahead of earnings announcements.² Consider, for instance, a firm like Apple. To profitably trade ahead of Apple's earnings announcements, an investor would need to collect information spanning multiple business segments and geographies. To the extent that this effort is costly,

²In the cross-sectional regressions, I account for firm size by directly controlling for lagged market capitalization (Ben-David et al., 2018) and by value-weighting observations. In the Appendix, I show my results are robust to instead including fixed-effects for deciles of market capitalization, formed each quarter.

Apple might have lower pre-earnings announcement price informativeness that is unrelated to the composition of its owners.

To establish a tighter causal link between passive ownership and pre-earnings announcement price informativeness, I build instruments for my baseline panel regressions using changes in passive ownership due to Russell 1000/2000 rebalancing (Appel et al. (2016), Ben-David et al. (2018), Gloßner (2018), Coles et al. (2022)) and S&P 500 index additions (Qin and Singal (2015), Bennett et al. (2020b)). My underlying assumption is that index rebalancing only affects price informativeness ahead of earnings announcements through its mechanical effect on passive ownership. Following Coles et al. (2022), I attempt to enforce this assumption by choosing an appropriate set of similar control firms that did not switch indices. For stocks switching to the Russell 2000, I choose a set of control firms which stayed in the Russell 1000 but were near the size cutoff used to determine index membership. I apply a similar logic for firms added to the S&P 500.

The IV estimates using both Russell and S&P 500 rebalancing reinforce a negative causal effect of passive ownership on pre-earnings announcement price informativeness. For example, when using the Russell rebalancing, I find that moving from the 90th percentile of passive ownership to the 10th percentile of passive ownership in 2018 decreases return volatility ahead of earnings announcements by 22.83 percentage points. Importantly, in the presence of the reverse causality described above, we would expect the OLS estimates to be biased upward in magnitude. The fact that the IV estimates are larger than those from the OLS suggest that the latter are not materially biased by these endogeneity concerns.

My preferred interpretation of these findings is that passive ownership decreases pre-earnings announcement price informativeness because passive investors gather less firm-specific information. This is motivated by recent work which has highlighted a trade off faced by passive investors, specifically that they may tend to increase the incorporation of systematic information at the expense of security-specific information (Cong et al. (2020), Glosten et al. (2021)). There are several pieces of evidence that lead me to favor this interpretation.

First, to show that the negative relationship between passive ownership and pre-earnings announcement price informativeness is being driven by firm-specific information, I examine how stock prices respond to fundamental news of a given size. The intuition is that if investors are not gathering private information before earnings announcements, they will have less

precise beliefs. When the news arrives, therefore, they will update their beliefs significantly, which leads to larger average price changes (Ganuza and Penalva (2010), Ganuza and Penalva (2010)). If passive investors are neglecting firm-specific information, this effect should be especially strong for the idiosyncratic component of earnings news.

To quantify how responsive stock prices are to earnings news, I run regressions in the spirit of Kothari and Sloan (1992), with earnings-day-returns on the left-hand side and standardized unexpected earnings (SUE) on the right-hand side. My regression estimates imply that a stock in the 90th percentile of passive ownership in 2018 responds nearly 3 times as much to earnings news as a stock in the 10th percentile of passive ownership. To refine this test, following Glosten et al. (2021), I decompose earnings news into systematic and idiosyncratic components. Consistent with my proposed mechanism, high passive stocks' increased responsiveness to earnings information is concentrated in the idiosyncratic component of news.

Next, I show direct evidence of differences in information collection between high and low passive stocks. In the cross-section, passive ownership is correlated with fewer Bloomberg terminal searches, evidence of less attention by institutional investors (Ben-Rephael et al., 2017). Specifically, moving from the 10th to the 90th percentile of passive ownership in 2018 is associated with 34% less abnormal institutional investor attention, relative to its whole-sample mean. Passive ownership is also correlated with fewer non-robot downloads of SEC filings, evidence of less fundamental research Loughran and McDonald (2017).³

I then discuss the effect of passive ownership on the supply of information, which I measure using data on sell-side analysts (Martineau and Zoican, 2021). The logic is that if information is costly to produce, sell-side analysts may respond to passive owners' decreased demand for information by supplying less or lower quality forecasts. Consistent with this hypothesis, passive ownership is correlated with decreased coverage, increased dispersion of analysts' estimates, decreased forecast accuracy and fewer updates.

Next, I examine trading volume before earnings announcements. This is useful to distinguish between theories that relate private information gathering to trading volume. Private information is a source of disagreement, which tends to increase trading volume (Wang, 1994), but it is also a source of information asymmetry, which can decrease trading volume

³These results are similar to those in Israeli et al. (2017) and Coles et al. (2022), who also provide evidence that less information is gathered about stocks with more passive ownership.

through fear of adverse selection (Foster and Viswanathan, 1990). Over the past 30 years, average abnormal pre-earnings volume has declined by about 5% relative to its whole-sample mean. In the cross-section, passive ownership is negatively correlated with pre-earnings trading volume, suggesting that the effect of reduced private information on disagreement dominates its effect on adverse selection.

Finally, I turn to the question of why the remaining non-passive investors don't fully offset the behavior of passive investors by gathering more information (Coles et al., 2022). My first proposed explanation is based on the finding in Ben-David et al. (2018) that ETFs (but not index mutual funds) increase non-fundamental volatility. I argue that this may create noise-trader risk (De Long et al., 1990), which could deter informed investors from learning about high passive stocks. Consistent with this, I show the effect of passive ownership on pre-earnings announcement price informativeness is stronger for ETFs than index mutual funds. My second explanation is that passive ownership may reduce pre-earnings liquidity, which is corroborated by my results on its relationship to pre-earnings turnover.

Overall, my analysis contributes to several strands of research on passive ownership and price informativeness. First, there have been mixed empirical results on the relationship between passive ownership and price informativeness. Part of this is due to the fact that, because information is hard to measure, prior work has relied on model-based measures of price informativeness. Motivated by different theoretical models, researchers have measured price informativeness in different ways and come to different conclusions.⁴ Using earnings announcements as a laboratory, I sidestep the need for a model-based measure of price informativeness, instead relying only on the assumption that earnings information is incorporated into prices quickly after it is released.

Focusing on earnings days, I find that there has been a trend toward decreased pre-announcement price informativeness over the past 30 years. Through cross-sectional regressions and two instrumental variables designs, I show passive ownership causes pre-earnings price informativeness to decline. In terms of magnitudes, averaging the point estimates from the OLS and both IVs implies that a 15% increase in passive ownership decreases QVS by

⁴For instance, Kacperczyk et al. (2018a) find a positive relationship when measuring price informativeness using the ability of current prices to forecast future fundamentals. Their approach is based on the noisy rational expectations models of Grossman and Stiglitz (1980) and Bai et al. (2016). In contrast, Bennett et al. (2020b) build on Roll (1988) and find a negative relationship when measuring price informativeness based on a regression of individual security returns on market-wide returns.

14.87, an approximately 16% decline relative to its mean in 1990 ($14.87/92.1 \approx 0.16$). Finally, I provide evidence for why passive ownership decreases price informativeness, specifically that passive investors gather less firm-specific information.

Literature Review. My paper contributes to a growing literature that studies the relationship between passive ownership and price informativeness. The conclusions from this research are mixed. Some studies find a positive link (Buss and Sundaresan (2020), Ernst (2020), Malikov (2020), Lee (2020), Kacperczyk et al. (2018a)), while others find a negative (Qin and Singal (2015), DeLisle et al. (2017), (Bond and Garcia, 2018), Garleanu and Pedersen (2018), Kacperczyk et al. (2018b), Breugem and Buss (2019), Brogaard et al. (2019), Bennett et al. (2020a), Bennett et al. (2020b)) or non-existent link (Coles et al., 2022). Part of the reason for this disagreement is that the papers differ in how they measure price informativeness. Another reason is that passive investors collect different types of information. For example, passive ownership may increase informativeness about systematic information while decreasing the incorporation of idiosyncratic information (Bhattacharya and O'Hara (2018), Cong et al. (2020), Antoniou et al. (2020), Glosten et al. (2021)).

The main difference between my approach and previous work is how I measure price informativeness. I focus specifically on the narrow window ahead of earnings announcements and measure how much of the news is incorporated into prices ahead of time. This allows me to abstract away from any particular model of price informativeness and only requires the assumption that prices reflect all of the information contained in the announcement shortly after its release. My main finding is that passive ownership causes less of the earnings information to be incorporated into prices ahead of the announcement. Moreover, I find that the effect occurs because passive owners gather less firm-specific information, consistent with theoretical work by Cong et al. (2020).

A related literature asks how price informativeness has evolved through time (Bai et al. (2016), Dávila and Parlato (2018)). These studies measure price informativeness by looking at time trends in the relationship between current prices and future fundamentals. The main conclusion from this work is that the link between prices and future fundamentals has become stronger over time, likely due to improvements in financial and information technology (Farboodi and Veldkamp (2020), Farboodi et al. (2020)). My analysis focuses on a different question, namely when information is incorporated into prices. I find that over time, there has been a trend toward a larger share of earnings information being incorporated

into prices after the news is released.⁵

2 Measurement & data

This section motivates my two measures of pre-earnings announcement price informativeness. I then describe the data I use to compute these measures and the firm-level passive ownership share. Finally, I present facts on the time-series decline in average pre-earnings announcement price informativeness and increase in passive ownership from 1990 to 2018.

2.1 Measurement

Ball and Brown (1968) show that prices incorporate a substantial portion of earnings news before it is actually made public. In the Appendix, Figure A.1 replicates their main finding, showing that prices increase before the release of good earnings news and drift down before the release of bad earnings news. A natural measure of pre-earnings announcement price informativeness, therefore, is the percentage of the total information which was incorporated into prices ahead of time.

I leverage this intuition to create two measures of pre-earnings announcement price informativeness. The first, which directly builds on the logic of Ball and Brown (1968), captures the share of total returns in the month leading up to and including an earnings announcement that occurs before the actual release. To this end, I define the pre-earnings drift magnitude (DM) for firm i with an earnings announcement at time t as:

⁵My results are not inconsistent with Bai et al. (2016), who show that current valuation ratios have become better predictors of long-horizon future cashflows. By the logic of Campbell and Shiller (1988), this pattern must be driven by the fact that valuation ratios covary less with future returns (Cohen et al., 2003). My results speak to something different, namely that there has been a change in *when* return volatility occurs. The time-series trends in DM and QVS show that more return volatility occurs after the release of earnings information. These trends say nothing about total return volatility per se and thus the covariance of valuation ratios with long-run future returns. So, it can both be true that valuation ratios have become better forecasts of long-run future earnings but in the short-run, prices anticipate earnings announcement news less. This could be the case, for example, if improvements in financial and information technology have led prices to better reflect earnings information after the news is released, as Bai et al. (2016) are using prices after the announcement of December calendar quarter earnings (i.e., prices from the end of March) to forecast future fundamentals. The same logic applies as to why my results are not inconsistent with those in Dávila and Parlato (2018).

$$DM_{i,t} = 100 \times \begin{cases} \frac{R_{i,(t-22,t-1)}}{R_{i,(t-22,t)}} & \text{if } r_{i,t} > 0 \\ \frac{R_{i,(t-22,t)}}{R_{i,(t-22,t-1)}} & \text{if } r_{i,t} < 0 \end{cases} = 100 \times \begin{cases} \frac{1}{1+r_{i,t}} & \text{if } r_{i,t} > 0 \\ 1 + r_{i,t} & \text{if } r_{i,t} < 0 \end{cases} \quad (1)$$

where $R_{i,(t-k,t+j)}$ denotes a cumulative *gross* market-adjusted return from $t - k$ to $t + j$ and $r_{i,t}$ denotes a *net* market-adjusted return, defined as the difference between firm i 's return and the market factor from Ken French's data library (Campbell et al., 2001). The choice of 22 trading-days (roughly a calendar month) before the announcement is in line with previous literature on pre-earnings price informativeness (Weller, 2018).⁶ When $r_{i,t}$ is positive, DM captures the percentage of the total gross return from $t - 22$ to t which is earned before the announcement itself. If $r_{i,t}$ is negative, this relationship would be reversed, which is why the measure is inverted when $r_{i,t}$ is less than zero. DM is designed to capture the share of earnings information that is incorporated into prices before the announcement.⁷ The assumption underlying this interpretation is that the earnings announcement is fully incorporated into prices quickly after its release (within a day). In this case, the return over the month before and through the announcement can be used to proxy for the total amount of information contained in the announcement. This interpretation of DM yields the first empirical prediction I use to measure the effect of passive ownership on price informativeness.

Prediction 1: *If passive ownership decreases pre-earnings announcement price infor-*

⁶In the Appendix, I show that my results are robust to including various post-earnings announcement windows in both my measures of pre-earnings announcement price informativeness.

⁷One concern with DM is that it does not depend on the pre-earnings announcement return, and rather only depends on $r_{i,t}$. For example, consider two stocks that have a 10% earnings-day return. Stock A has a cumulative pre-earnings announcement return ($r_{i,(t-22,t-1)}$) of 0% while stock B has a cumulative pre-earnings announcement return of 100%. Both of these stocks have the same DM , which seems counter-intuitive given that the B's earnings-day return is smaller than A's, *relative to its pre-earnings announcement run-up*. This concern is addressed in my second measure of pre-earnings announcement price informativeness, QVS , which compares the volatility of earnings-day returns to pre-earnings announcement volatility.

An alternative is to work with net returns instead of gross returns and define $\widetilde{DM} = r_{i,(t-22,t-1)}/r_{i,(t-22,t)}$. In this example, stock A would have a \widetilde{DM} of 0 while stock B would have a \widetilde{DM} of 0.83, consistent with the intuition that more information was incorporated into stock B's price ahead of the earnings announcement. One issue with using net returns is that \widetilde{DM} is not well defined when $r_{i,(t-22,t)}$ is close to zero. Weller (2018), who uses a measure of pre-earnings announcement price informativeness similar to \widetilde{DM} , solves this by removing "non-events" i.e., earnings announcements where the total cumulative return is near zero. This filter, however, removes the majority of earnings announcements in his sample (54.5%). DM has the advantage that it can be computed for every earnings announcement. Another issue with \widetilde{DM} is that it will not be responsive to changes in post-earnings announcement returns if the pre-earnings announcement return is close to zero. This is not a problem for DM , which is always responsive to changes in $r_{i,t}$.

mativeness, it should cause DM to decline

Using DM to measure informativeness does, however, have several limitations. For example, consider two firms with cumulative returns of 0% over the pre-earnings announcement month. One of them has an earnings day return of 5%, while the other has an earnings day return of -5%. Intuition suggests that these two firms have equally informative pre-earnings announcement prices, but they will have slightly different values of $DM_{i,t}$ (95.24 vs. 95.00).

The second issue is that DM is sensitive to the level of volatility. To fix ideas, consider two stocks with different volatilities: Leading up to an earnings announcement, Stock A has alternating returns of $\pm 1\%$ while stock B has alternating returns of $\pm 5\%$. On the announcement day, stock A has a return of 1% while stock B has a return of 5%. It seems natural that both stocks have equally informative pre-earnings announcement prices, as the earnings day returns are the same magnitude as those over the prior month. These stocks will, however, have significantly different values of DM (99.01 vs. 95.24).

To address these limitations, I build on the model of information in Ganuza and Penalva (2010) and create a measure based on the share of total volatility which occurs before the earnings announcement. Specifically, I define the quadratic variation share (QVS) for firm i around earnings announcement t as:

$$QVS_{i,t} = 100 \times \sum_{\tau=-22}^{-1} r_{i,t+\tau}^2 / \sum_{\tau=-22}^0 r_{i,t+\tau}^2 \quad (2)$$

where $r_{i,t}$ denotes a market-adjusted daily return.⁸

QVS measures the contribution of the earnings day return to volatility in the month leading up to and including the earnings announcement.⁹ Like DM , if more of the information contained in a given earnings announcement is being incorporated into prices ahead of the release, the magnitude of the earnings day return should be smaller, and so too will QVS . This interpretation of QVS yields the second empirical prediction I use to measure

⁸All my results are robust to defining QVS at the annual level i.e., defining the numerator to be the sum of squared non-earnings returns in year t , while the denominator is the sum of squared returns for all days in year t . This is evidence that my results are not driven by the choice of a 22-day pre-earnings window.

⁹Recent work by Dávila and Parlato (2021) has shown that the relationship between volatility and price informativeness can be non-monotonic. This raises the question of whether QVS is sensitive to differences in the overall level of volatility across stocks and time. In Appendix F, I simulate a Kyle (1985)-style model and show that QVS is robust to changes in the level of fundamental volatility or the intensity of noise-trading activity.

the relationship between passive ownership and price informativeness.

Prediction 2: *If passive ownership decreases pre-earnings announcement price informativeness, it should cause $QVS_{i,t}$ to decline*

There are many other ways to measure price informativeness. For example, Weller (2018) quantifies pre-earnings announcement price informativeness using a measure similar to DM , defined using net returns instead of gross returns. I discuss these alternatives and show my conclusions are robust to them in Appendix C.

2.2 Data

My sample starts with all ordinary common shares (share codes 10-11) traded on major exchanges (exchange codes 1-3) that can be matched between CRSP and IBES between 1990 and 2018. For each stock, around each earnings announcement, I need to construct DM , QVS and the level of passive ownership.

To construct the measures of price informativeness, I need to identify the first time investors could have traded on earnings information during normal market hours. I identify these days using the earnings release date and time in IBES. 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 in CRSP is labeled as the effective earnings date. To be included in the final sample, a firm must have non-missing returns in CRSP each day from $t - 22$ to t around the earnings announcement. I use these returns to construct DM and QVS .

The last object I need to construct for each observation is passive ownership, which I define as the fraction of a stock's shares outstanding which are held by passive funds. Following Appel et al. (2016), I identify passive funds using the CRSP mutual fund database, selecting all index funds, all ETFs and all funds with names that identify them as index funds. To calculate how many shares of each stock passive funds hold, I use the WRDS MFLINKS database to match the identified funds to Thompson S12, which contains data on funds' holdings. The passive ownership share is the sum of all shares held by passive funds, divided by shares outstanding in CRSP. In the Appendix, I show that my results are quantitatively unchanged by dropping all observations with zero passive ownership (Dannhauser, 2017).

2.3 Basic properties

To visualize the time-series and cross-sectional properties of the three key variables in my analysis, Figure 1 plots the 25th percentile, median, 75th percentile and value-weighted average of QVS , DM and passive ownership. The top left panel shows that QVS decreased steadily over my sample. Average QVS decreased from 92.1% in 1990 to 72.4% in 2018. This 19.7 percentage point decline is about the same size as QVS 's whole-sample standard deviation of 21.1%.¹⁰ There has also been a trend toward increased cross-sectional spread in QVS , with the interquartile range increasing from 10% to over 40%.

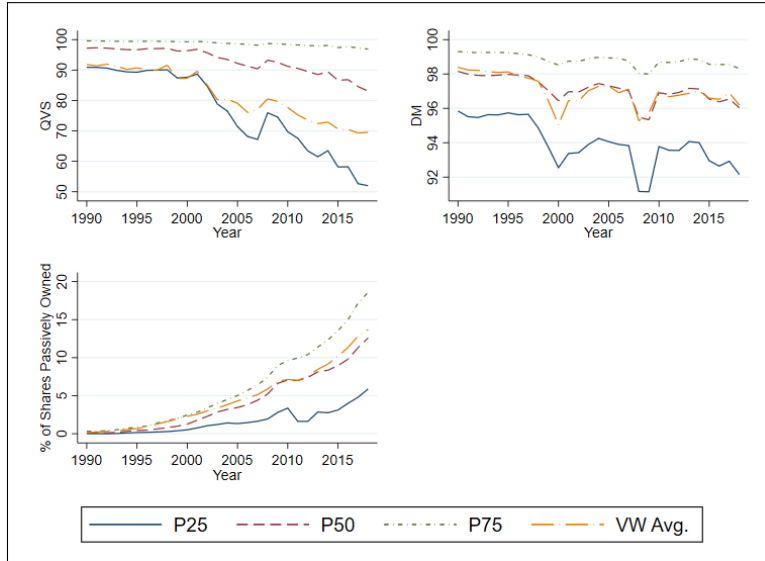


Figure 1. Trends in QVS , DM and passive ownership, 1990-2018. To compute the value-weighted average ($VW Avg.$), within each quarter, observations are weighted in proportion to their market capitalization at the end of the previous quarter. DM and QVS are defined in Equations 1 and 2. Passive ownership is defined as the fraction of a stock's shares which are held by all index funds, all ETFs and all mutual funds with names that identify them as index funds.

The decline in average QVS accelerates around 2001, which coincides with two changes to the amount of information released before earnings announcements. The first is Regulation Fair Disclosure (Reg FD), passed in August 2000, which reduced early selective disclosure of earnings information. The second is the increased enforcement of insider trading laws (Coffee, 2007).

¹⁰The Appendix shows that the decrease in QVS was due to a simultaneous increase earnings-day volatility and a decrease in non-earnings-day volatility.

The top right panel of Figure 1 shows that, consistent with the trend in *QVS*, average *DM* decreased by about 2 between 1990 and 2018. This drop is roughly 40% of *DM*'s whole sample standard deviation of 5.4. There are notable drops in average *DM* in the early 2000s and again in the late 2000s. As with *QVS*, the level shift down in the early 2000s may be the result of Reg FD and decreased insider trading.

Another explanation for the drop in *DM* is that these years correspond to the dot-com boom and the Global Financial Crisis. These were periods with higher overall volatility, leading to larger absolute earnings-day returns and lower values of *DM* on average. *QVS* may not have experienced a correspondingly large drop in the dot-com boom because it explicitly accounts for the level of volatility in the month leading up to the earnings announcement.

The bottom left panel of Figure 1 shows that passive ownership steadily increased over my sample. From 1990 to 2018, average passive ownership went from nearly zero to owning almost 15% of the US stock market. These numbers closely mirror those in the ICI factbook. Like *QVS* and *DM*, the difference between high and low passive ownership stocks also grew over my sample, with the interquartile range increasing from 0% in 1990 to about 15% by 2018. Table 1 contains summary statistics on the measures of pre-earnings announcement price informativeness measures, as well as passive ownership.

		25%	50%	Mean	75%	St. Dev.
<i>QVS</i>		89.82	96.97	91.35	99.46	14.06
<i>DM</i>	1990-1999	95.36	97.84	96.41	99.17	4.48
<i>Passive</i>		0.06	0.38	0.69	1.03	0.82
<i>QVS</i>		61.23	88.37	76.68	97.91	26.19
<i>DM</i>	2010-2018	93.32	96.77	95.04	98.68	5.48
<i>Passive</i>		3.21	8.37	8.75	12.83	6.55
<i>QVS</i>		80.64	94.87	85.17	99.06	21.11
<i>DM</i>	All Years	94.08	97.26	95.50	98.92	5.36
<i>Passive</i>		0.32	1.66	3.69	5.18	4.90

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

3 Passive ownership and pre-earnings announcement price informativeness in the cross-section

This section documents the relationship between passive ownership and pre-earnings announcement price informativeness. It starts with cross-sectional regressions of QVS and DM on passive ownership. Across both measures, the regressions show that higher passive ownership is correlated with decreased pre-earnings announcement price informativeness. I then provide evidence that options markets internalize the relationship between passive ownership and earnings-day volatility. Finally, I perform robustness checks to show that Regulation Fair Disclosure and the rise of algorithmic trading are not driving my OLS regression estimates.

3.1 Baseline analysis

I run the following regression to measure the relationship between pre-earnings announcement price informativeness and passive ownership:

$$\text{Price informativeness}_{i,t} = \alpha + \beta \text{Passive}_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t} \quad (3)$$

where $\text{Price informativeness}_{i,t}$ is either $QVS_{i,t}$ or $DM_{i,t}$. Controls in $X_{i,t}$ include time since listing (age), one-month lagged market capitalization, returns from month $t-12$ to $t-2$, one-month lagged book-to-market ratio and the institutional ownership ratio.¹¹ $X_{i,t}$ also includes CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility, all computed over the previous 252 trading days. The controls in $X_{i,t}$ are selected to capture firm characteristics known to be correlated with passive ownership (Glosten et al., 2021). The Appendix contains details on the construction of all control variables.

Equation 3 also includes firm and year-quarter fixed effects. The firm fixed effects account for differences in average price informativeness e.g., investors may pay more attention to Apple’s earnings announcements than to those of Dominion Energy. The year-quarter fixed effects account for the time trends in pre-earnings announcement price informativeness and

¹¹It’s possible that firm size has a non-linear effect on price informativeness. In the Appendix, I show the cross-sectional regression results are quantitatively unchanged by including fixed effects for deciles of market capitalization, formed each quarter.

the seasonality in earnings news. Standard errors are double-clustered at the firm and year-quarter level.

The regression results are in Table 2. Consistent with prediction 2, Column 1 shows that there is a negative relationship between passive ownership and QVS . The point estimate implies that a firm in the 90th percentile of passive ownership in 2018 (25%) has 9.1 pp lower QVS than a firm in the 10th percentile of passive ownership in 2018 (2%). For reference, 9.1 is roughly $2/5^{th}$ of QVS 's whole sample standard deviation. To allay concerns that small firms are driving my results, Column 2 weights observations by each firm's share of total market capitalization at the end of the previous quarter. Using value weights shrinks the estimated coefficient, but it remains statistically significant at the 1% level.

Column 3 shows that, consistent with prediction 1, there is also a negative correlation between DM and passive ownership. The point estimate implies that a firm in the 90th percentile of passive ownership in 2018 has 1.05 lower DM than a firm in the 10th percentile of passive ownership in 2018. For reference, 1.05 is approximately $1/5^{th}$ of DM 's whole sample standard deviation. Column 4 shows that the relationship between passive ownership and DM is quantitatively unchanged by value weighting observations.

3.2 Evidence from options

Columns 1 and 2 of Table 2 show that stocks with more passive ownership have relatively more volatility on earnings announcement days. If options markets internalize this relationship, then we would expect options exposed to earnings announcement risk to be relatively more expensive for high passive stocks. To quantify this effect, I adapt Kelly et al. (2016)'s Implied Volatility Difference (IVD) to measure how much higher implied volatility is for options that span earnings announcements, relative to options that expire the month before and after the announcement.

Specifically, letting τ denote an earnings announcement, I identify regular monthly expiration dates a , b and c , such that $a < \tau < b < c$. Then, I calculate the average implied volatility (in percentage points) \overline{IV}_i for at the money options on stock i with these expiration dates. The final variable of interest, the implied volatility difference, is defined as:

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2} (\overline{IV}_{i,a} + \overline{IV}_{i,c}) \quad (4)$$

	QVS		DM	
	(1)	(2)	(3)	(4)
Passive Ownership	-39.48*** (3.06)	-26.23*** (9.72)	-4.78*** (0.61)	-4.96*** (1.25)
Observations	430,489	430,489	430,489	430,489
R-Squared	0.23	0.24	0.22	0.28
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value

Table 2 Cross-sectional regression of price informativeness on passive ownership. Table with estimates of β from:

$$\text{Price informativeness}_{i,t} = \alpha + \beta \text{Passive}_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$$

where Price informativeness $_{i,t}$ is either $QVS_{i,t}$ or $DM_{i,t}$. Controls in $X_{i,t}$ include 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.

where higher values of $IVD_{i,\tau}$ imply that options which span earnings announcements are relatively more expensive than those not exposed to earnings announcement risk.¹² The sample for IVD is shorter than for QVS and DM because it relies on OptionMetrics, which begins in 1996.

My preferred interpretation of IVD is built on the same logic as QVS . As fewer investors gather information ahead of earnings announcements, volatility on the announcement day itself should increase (Ganuza and Penalva (2010), Åstebro and Penalva (2022)). If options markets internalize the negative relationship between passive ownership and pre-earnings announcement information gathering, the associated effect on earnings-day volatility should be reflected in higher option prices. This interpretation yields a testable prediction for the relationship between IVD and passive ownership.

¹²See the Appendix for step-by-step details on how I construct IVD . 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. This concern is partially alleviated by the inclusion of firm fixed effects. In addition, all the results are qualitatively unchanged instead defining the implied volatility difference as a ratio: $\overline{IVD}_{i,\tau} = \overline{IV}_{i,b} / \frac{1}{2} (\overline{IV}_{i,a} + \overline{IV}_{i,c})$.

Prediction 3: *If passive ownership decreases pre-earnings announcement information gathering, it should cause IVD to increase*

In terms of basic properties, the Appendix shows that average *IVD* is positive and has increased by about 5 percentage points over the past 25 years. This is consistent with the decline of *QVS* and *DM*, suggesting the trends towards decreased pre-earnings announcement price informativeness in Figure 1 are reflected in option prices.

To test prediction 3, I run a regression of *IVD* on passive ownership and the same controls and fixed effects as Equation 3. Table 3 contains the results. Column 1 shows that, consistent with prediction 3, *IVD* is positively correlated with passive ownership. The estimated coefficient implies that a firm in the 90th percentile of passive ownership in 2018 has a 2.16 percentage point higher average *IVD* than a firm in the 10th percentile of passive ownership in 2018. For reference, the whole sample mean of *IVD* is 5.07, and its standard deviation is 9.37. Column 2 shows the relationship between passive ownership and *IVD* is even stronger when using value weights instead of equal weights. These results imply that options markets reflect the relationship between passive ownership and pre-earnings announcement price informativeness, corroborating the findings in Table 2.

3.3 Additional robustness

One threat to my OLS regression results is Regulation Fair Disclosure (Reg FD), passed in August 2000, which reduced the early release of earnings information. Even though all the specifications in Table 2 have time fixed effects, this threat remains because Reg FD may have differently affected stocks with more passive ownership. Columns 3 and 4 of Table 3 show that the OLS estimates are qualitatively unchanged when using only earnings announcements between 2001 and 2018, evidence that Reg FD is not driving my results.

Another threat to my OLS regressions is the rise of algorithmic trading (AT), which can reduce the returns to informed trading (Weller, 2018). This could threaten my results if e.g., high passive stocks also have high AT activity due to ETF arbitrage. Columns 5 and 6 replicate the baseline regressions, but explicitly control for the AT measures in Weller (2018).¹³ The OLS estimates are not significantly changed by including these controls,

¹³These AT 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 OLS regressions. See the Appendix for a detailed description of how the AT measures are constructed.

	IVD		Post Reg FD		AT Controls	
	(1)	(2)	QVS (3)	DM (4)	QVS (5)	DM (6)
Passive Ownership	9.81*** (2.65)	17.27*** (3.26)	-33.59*** (3.52)	-4.36*** (0.71)	-28.25*** (5.05)	-4.918*** (1.13)
Observations	111,415	111,415	250,068	250,068	80,990	80,990
R-Squared	0.35	0.45	0.22	0.22	0.25	0.27
Firm + Year/Quarter FE	✓	✓	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓	✓	✓
Weight	Equal	Value	Equal	Equal	Equal	Equal

Table 3 Corroborating evidence for effect of passive ownership on pre-earnings announcement price informativeness.

Estimates of β from:

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

In Columns 1-2, the left-hand side variable is *IVD*, while in Columns 3 and 5 it is *QVS* and in Columns 4 and 6 it is *DM*. Controls in $X_{i,t}$ include 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 and firm fixed effects. Columns 3-4 restrict to observations between 2001 and 2018. Columns 5-6 restrict to observations that can be matched to the SEC MIDAS data, and also include controls for the AT measures in Weller (2018). Standard errors double clustered at the firm and year-quarter level in parenthesis.

evidence that a correlation between AT activity and passive ownership is not driving my results.

4 Causal evidence

One limitation of the regressions in Table 2 is that passive ownership is not randomly assigned in the cross-section of stocks. It's possible, therefore, that passive ownership increased the most in stocks with low pre-earnings announcement price informativeness and causality runs the other way. For example, Figure D.1 in the Appendix shows that passive ownership has a strong positive correlation with market capitalization. Large firms may be harder to value, because e.g., they are made up of multiple business segments (Cohen and Lou, 2012). In this case, we might expect large firms to have lower pre-earnings announcement price informativeness for reasons unrelated to their larger passive ownership share.

In my setting, reverse causality seems unlikely because a significant amount of passive ownership is determined by mechanical rules e.g., being one of the 100 lowest volatility stocks in the S&P 500 (Invesco’s S&P 500 Low Volatility ETF, SPLV) or having one of the 1000 largest float-adjusted market capitalizations in the Russell 3000 (iShares’ Russell 1000 ETF, IWB). Ex-ante, it’s not obvious why the intersection of these rules would select stocks with low pre-earnings announcement price informativeness.

Even so, the cross-sectional correlations do not conclusively establish a causal link between passive ownership and pre-earnings announcement price informativeness. To establish causality, I construct two instruments for passive ownership using changes in index membership due to Russell 1000/2000 rebalancing and S&P 500 additions. Both IV designs are built on the logic of difference-in-differences. To this end, I identify a group of treated firms that experience a mechanical increase in passive ownership due to an index change. Then, to alleviate concerns of selection bias, I identify a corresponding group of similar control firms that do not. Finally, I instrument for passive ownership using the expected change in passive ownership from switching indices. My IV estimates confirm a negative causal relationship between passive ownership and pre-earnings announcement price informativeness.

4.1 Identifying treated & control firms

Until 2006, at the end of each May, FTSE Russell selected the 1000 largest stocks by float-adjusted market capitalization to be members of the Russell 1000, and selected the next 2000 largest stocks to be members of the Russell 2000. To reduce turnover between the two indices, in 2007, Russell switched to a banding rule. Now, as long as a potential switcher’s market capitalization is within $\pm 2.5\%$ of the Russell 3000E’s total market capitalization, relative the 1000th ranked stock (the upper and lower bands), it will remain in the same index as the previous year.

Moving from the 1000 to the 2000 mechanically increases the fraction of a firm’s shares that need to be held by passive funds. One reason is that switchers go from being one of the smallest stocks in a value-weighted index of big stocks, to one of the biggest stocks in a value-weighted index of small stocks, significantly boosting their index weight (Appel et al., 2016). Another reason is that the Russell 2000 has a higher average passive ownership share than the Russell 1000 (Pavlova and Sikorskaya, 2022).

In this setting, the ideal difference-in-differences design would compare potential switchers to those that actually switched. Identifying possible switchers is not straightforward, however, as the data that Russell uses to compute May market capitalizations is not made available to researchers. To compute a proxy for the Russell May market capitalizations, I follow the method in Coles et al. (2022).¹⁴ Using their May market capitalization proxy, I correctly predict Russell 1000/2000 index membership for 98.63% of Russell 3000 stocks in my sample.

I also follow Coles et al. (2022) to identify groups of treated and control firms. Each May, I create a cohort of possible switchers that were in the Russell 1000 the previous year. From 1990-2006, this is firms within ± 100 ranks around the 1000th ranked stock, while from 2007-2018, this is firms within ± 100 ranks of the lower band. The treated firms are those that ended up switching, while the control firms are those that stayed in the 1000.¹⁵ 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 are near the index assignment threshold in multiple years, but don't switch.¹⁶ These filters yield about 700 treated firms and 600 control firms.

My second set of treated and control firms are built using additions to the S&P 500. 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, it experiences an increase in passive ownership, as the index mutual funds and ETFs tracking the index need to buy the stock.¹⁷

¹⁴I would like to thank the authors for sharing their replication code with me. The Appendix contains a step-by-step explanation of how I compute the May market capitalization proxy.

¹⁵Another natural set of treated firms are those that switch from the Russell 2000 to the Russell 1000 because they experience a decrease in passive ownership. In the Appendix, I show that within one year of switching, this decrease is totally offset by the time trend toward increased passive ownership.

¹⁶One concern with defining treatment as switching to the 2000 instead of switching to *and staying in* the 2000 is that firms may change their index status in the post-treatment period. One could instead require treated firms to be out of the 2000 for the whole pre-treatment period and in the 2000 for the entire post-treatment period. This, however, is not my preferred specification, as whether or not a firm stays in/out of a particular index is endogenous and future index status is not known at the time of index addition.

¹⁷A natural extension is to examine firms that are dropped from the S&P 500 index, which experience a decrease in passive ownership. As I discuss in the Appendix, this is a less ideal setting than index addition, as firms are usually dropped from the index for (1) poor performance or lack of liquidity, which is related to firm fundamentals or (2) being acquired by or merged with another firm in which case there will be no post-index-deletion observations.

One concern with defining treatment as being added to the S&P 500 is that these changes are determined by a committee, rather than a mechanical rule. Therefore, it's possible that the increase in passive ownership is not fully exogenous to firm fundamentals. To ameliorate this concern, I follow the logic in the previous subsection and carefully choose a set of comparable control firms.

I start by obtaining 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, form quintiles of market capitalization. For each added firm, the first set of control firms are those in the same three-digit SIC industry and same quintile of industry market capitalization which are outside the S&P 500 index. I also form a second control group of firms in the same 3-digit SIC industry and market capitalization quintile, but that are already in the S&P 500 index. Cohorts are defined as all matched treated and control firms in the same industry and size bucket in a given month.

As with the Russell 1000/2000 switchers, control firms can appear in more than one cohort. For example, the same firm outside the index can be a control for multiple firms added to the index at different points in time. These filters yield about 500 treated firms, 600 control firms in the index and 2,000 control firms out of the index.

4.2 Effect of treatment on passive ownership

The next step in building the IV is quantifying the effect of being treated on passive ownership (the first stage). To visualize this, the top left panel of Figure 2 compares the level of passive ownership around the index rebalancing month between Russell switchers and stayers. Within each cohort, I subtract the average level of passive ownership to ease comparison across years. Reassuringly, pre-addition changes and levels of passive ownership are similar between the treated and control groups. The treated firms, however, experience an increase in passive ownership at $t = 0$ and remain at a higher level of passive ownership over the next 12 months.¹⁸

The top right panel of Figure 2 shows the level of passive ownership for S&P 500 additions

¹⁸Russell reconstitutions always coincide exactly with the end of a calendar quarter, so Figure 2 only plots data points for months with S12 filings (the last month of each calendar quarter).

and matched control firms around the month of index rebalancing. Again, within each cohort, I subtract the average level of passive ownership to facilitate the comparison across industry-size buckets 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. After index addition, the added firms experience an increase in passive ownership, essentially going from the level of the control firms outside the index to the level of control firms inside the index.¹⁹

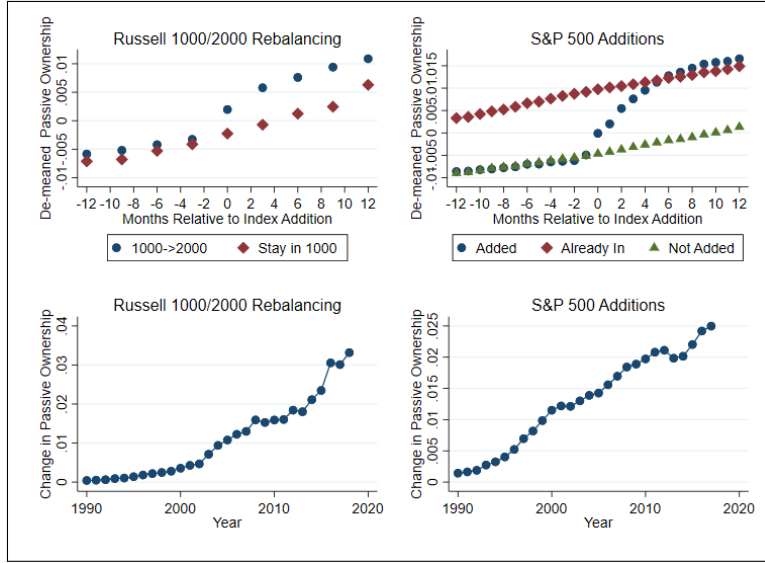


Figure 2. Effect of treatment on passive ownership. Top left panel: Average level of passive ownership for firms that stay in the Russell 1000 (“Stay in 1000”) and firms that switched from the Russell 1000 to the Russell 2000 (“1000 → 2000”). Top right panel: Average level of passive ownership for control firms out of the index (“Not Added”), control firms in the index (“Already In”) and added firms (“Added”). For both top panels, passive ownership is demeaned within each group of matched treated and control firms. Bottom left panel: 5-year moving average change in passive ownership for Russell 1000 to 2000 switchers from month $t = -3$ to $t = 3$ around the reconstitution date by year. Bottom right panel: 5-year moving average change in passive ownership for S&P 500 additions from month $t = -3$ to $t = 3$ around the index rebalancing date by year.

¹⁹S&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 always 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 2, between quarter ends, I fix passive ownership at its last reported level each month. This is why passive ownership appears to increase slowly around the month of index addition, as I am averaging across observations with differences in time until the first set of post-index-addition S12 filings are released.

As shown by Figure 1, aggregate passive ownership has been increasing over time. One consequence of this trend is that the increase in passive ownership associated with switching from the Russell 1000 to the Russell 2000 and being added to the S&P 500 has grown over my sample. The two bottom panels of Figure 2 show the average change in passive ownership for treated firms between month $t = -3$ and month $t = 3$ relative to the index reconstitution. For Russell 2000 switchers, the increase grew from almost nothing in 1990 to about 3.5% by 2018. The change in passive ownership accelerated after 2000, the year IWM (the largest Russell 2000 ETF) was launched. The change in passive ownership from being added to the S&P 500 exhibits a similar trend.

Given the trends in the bottom two panels of Figure 2, my IV design needs to account for the time series variation in passive ownership associated with index changes. To this end, I create a proxy for the expected increase in passive ownership from being treated, which I call $\text{Passive Gap}_{i,t}$. For the Russell switchers, it is defined as the difference in passive ownership between firms in the Russell 1000 and the Russell 2000 within ± 100 ranks of the 1000th ranked firm in March (the last S12 filing date before index rebalancing). For the S&P 500 additions, $\text{Passive Gap}_{i,t}$ 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 $\text{Passive Gap}_{i,t}$ for all other added firms that year.

4.3 Instrumental variables design

The logic behind my IV is to use being treated, the post-treatment period and $\text{Passive Gap}_{i,t}$ to instrument for passive ownership. The three key pieces of the IV are therefore: (1) the instrumented change in passive ownership (2) the IV specification and (3) the reduced form:

$$\text{Passive}_{i,t} = \alpha + \beta_1 \text{Post}_{i,t} + \beta_2 \text{Passive Gap}_{i,t} \times \text{Treated}_{i,t} \times \text{Post}_{i,t} + FE + \epsilon_{i,t} \quad (5)$$

$$\text{Outcome}_{i,t} = \alpha + \beta_3 \widehat{\text{Passive}_{i,t}} + FE + \epsilon_{i,t} \quad (6)$$

$$\text{Outcome}_{i,t} = \alpha + \beta_4 \text{Post}_{i,t} + \beta_5 \text{Passive Gap}_{i,t} \times \text{Treated}_{i,t} \times \text{Post}_{i,t} + FE + \epsilon_{i,t} \quad (7)$$

where $\text{Outcome}_{i,t}$ is *QVS* or *DM* and $\text{Post}_{i,t}$ is an indicator for observations after the index change. Following Coles et al. (2022), all three equations include firm-by-cohort fixed effects.

I restrict to data within three years before or after index addition, but exclude three months immediately before or after the event to avoid index inclusion effects (Morck and Yang (2001), Madhavan (2003)). $\text{Passive Gap}_{i,t} \times \text{Treated}_{i,t}$ is not included in the first stage or reduced form because it is constant within each firm-cohort and therefore is fully explained by the fixed effects. Standard errors are double clustered at the firm and quarter level.

Panel A of Table 4 shows the results of the IV built on Russell rebalancing and Column 1 shows the first stage. The associated F-statistic is large, which is not surprising given the increase in passive ownership pictured in Figure 2.²⁰ The coefficient on $\text{Post} \times \text{Treated} \times \text{PassiveGap}$ is larger than 1, implying that *PassiveGap* tends to understate the actual change in passive ownership associated with switching to the Russell 2000. One reason for this is that there are three years of post-rebalancing observations for the treated firms and the trend toward increased passive ownership has been steeper for Russell 2000 firms than Russell 1000 firms.

Column 2 is the instrumental variables (IV) specification with *QVS* on the left hand side. The effect of passive ownership on *QVS* is negative, consistent with the cross-sectional regression results. The IV estimate of -99.28 is about 2.5 times the OLS estimate of -39.48 . In Column 4, the analogue to Column 2 for *DM*, the IV estimate of -13.01 is also negative and about 2.5 times the OLS regression coefficient of -4.78 . Importantly, in the presence of the reverse causality described at the start of this section, we would expect the OLS estimates to be biased upward in magnitude. The fact that the IV estimates are more negative than the OLS estimates suggest that the latter are not materially biased by this endogeneity concern.

Although the reduced-form estimates for *QVS* and *DM* are the same sign as the OLS estimates, they are statistically insignificant. It is not obvious, however, that the reduced form estimates should be directly comparable with the OLS results. One reason is that the cross-sectional regressions use the *level* of passive ownership, while the reduced form uses the expected *change* in passive ownership from index changes (i.e., $\text{Passive Gap}_{i,t}$), which may only be informative about the sign of the treatment effect. I present a more detailed discussion of the differences between the IV and RF specifications in the next subsection and

²⁰Because I am using both *Post* and $\text{Post} \times \text{Treated} \times \text{Passive Gap}_{i,t}$ as instruments for passive ownership, the time trend and the change in passive ownership associated with being treated in Figure 2 are jointly driving the large magnitude of the F-statistic in Table 4. In the first stage, both *Post* and $\text{Post} \times \text{Treated} \times \text{Passive Gap}_{i,t}$ are individually statistically significant at the 1% level.

in the Appendix.

Panel A: Russell Rebalancing					
	First Stage	QVS IV	RF	DM IV	RF
Post	0.017*** (0.003)		0.45 (0.56)		0.16 (0.10)
Post x Treated x Passive Gap	4.727*** (0.297)		-668.05*** (64.01)		-108.56*** (12.86)
Passive Ownership		-136.40*** (13.23)		-21.57*** (2.62)	
Observations	31,032	31,032	31,032	31,032	31,032
F-statistic	133				
Panel B: S&P 500 Additions					
	First Stage	QVS IV	RF	DM IV	RF
Post	0.014*** (0.001)		-2.42*** (0.28)		-0.29*** (0.08)
Post x Treated x Passive Gap	0.547*** (0.044)		-24.48 (18.11)		-3.32 (4.33)
Passive Ownership		-158.17*** (17.36)		-18.96*** (5.37)	
Observations	185,494	185,494	185,494	185,494	185,494
F-statistic	388				
Cross-sectional regression estimate		-39.48	-39.48	-4.78	-4.78

Table 4 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness. Estimates from:

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

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

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

where $Outcome_{i,t}$ is QVS or DM and $Post_{i,t}$ is an indicator for observations after the index change. $Passive\ Gap_{i,t}$ is the expected change in passive ownership from being treated. Column 1 in each panel is a first-stage regression. Columns 2 and 4 are instrumental variables regressions. Columns 3 and 5 are reduced-form regressions. Panel A contains observations from Russell rebalancing, while Panel B contains observations from S&P 500 additions. FE are fixed effects for each cohort. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

Panel B of Table 4 is the analogue of Panel A using S&P 500 additions. Consistent with Panel A, the first stage regression in Column 1 has a large F-statistic. Columns 2 and 4 are the IV regressions, which both show a negative and statistically significant relationship between passive ownership and pre-earnings announcement price informativeness. Like Panel

A, these point estimates are larger in magnitude than the cross-sectional regression estimates by a factor of about 3. 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.²¹

4.4 Discussion

The assumption underlying my IV strategy is that index addition only affects price informativeness through its associated effect on passive ownership. One threat to this is that index switching/addition may be associated with an increase in total institutional ownership (Boone and White, 2015). Gloßner (2019) shows, however, that although there is an increase in passive ownership following Russell index reconstitution events, there is little change in overall institutional ownership.²² To further alleviate the concern that institutional ownership is driving my results, in the Appendix, I show the IV results are quantitatively unchanged by including the institutional ownership ratio on the right-hand side.

A second issue, as discussed above, is that in both panels of Table 4, the IV is always significant, while the reduced form is insignificant. The concern is that, as discussed in Chernozhukov and Hansen (2008a), a significant IV with an insignificant reduced form potentially indicates weak instruments. This is likely not a problem in my setting, as the first stage is very strong ($F > 200$ for the Russell 1000 to 2000 switchers and $F > 300$ for the S&P 500 additions). In the Appendix, following Lochner and Moretti (2004), I show why we might expect from a purely econometric perspective the reduced form to be less significant than the IV.²³ I also discuss how the presence of shadow indexing (Mauboussin, 2019) could

²¹Suppose that what truly matters for price informativeness is the total amount of passive ownership. My measure, $Passive_{i,t}$, only captures funds that are explicitly passive, and misses e.g., shadow index funds (Mauboussin, 2019), as well as institutions that do index replication internally. If firms added to the S&P 500 experience an increase in these types of non-explicit passive ownership as well, we might expect their price informativeness to decline more than would be explained by index fund holdings alone.

²²A related concern, raised in Appel et al. (2020), is that for the Russell switchers, the treatment is correlated with firm size. Given that my results are similar using both switching from the Russell 1000 to the Russell 2000, which applies to shrinking firms and S&P 500 index addition, which applies to growing firms, I find it unlikely that a pure size effect is driving my results.

²³The result in Lochner and Moretti (2004) regards the ratio between the IV and reduced form t-statistics. In an infinitely large sample, however, the IV and reduced form should yield the same conclusion, even if this ratio is large. In the Appendix, I show through simulations that in a finite sample with 30,000 observations (i.e., the size of the sample in Panel A of Table 4), it is possible for the IV to be significant but the reduced form to be insignificant.

explain why the IV is significant but the reduced form is not.

An additional concern with the results in Table 4 is that many previous studies have used switching between from the Russell 1000 to the Russell 2000 and additions to the S&P 500 as natural experiments when studying the effects of passive ownership on a variety of outcomes e.g., corporate governance, disclosure and investment. As discussed in Heath et al. (2020), this re-use of natural experiments can lead to false positives in later studies. The particular issue is that my results could be driven by the effects of passive ownership on previously documented outcomes, rather than passive ownership per se.

The solution proposed by Heath et al. (2020) is to use t-statistics which explicitly account for how many times the natural experiment has been re-used. Table 4 shows that almost all of my IV t-statistics are over 3.62. This implies that even if previous research had looked at the effect of these index changes on over 300 other distinct outcomes, my results are unlikely to be spurious. Further, the Russell switcher IV yields similar point estimates to the S&P addition IV, even though these index changes have different implications for other known outcomes (e.g., firm size), again allaying concerns that my results are driven by factors other than passive ownership.

Finally, to confirm the robustness of my causal estimates, in the Appendix I build on the logic in Bernstein (2015) to develop an alternative IV strategy. Specifically, I use the interaction between a firm’s CAPM beta at the end of March and cumulative market return from the start of April to the Russell ranking date in May to instrument for passive ownership over the year starting in July. Crucially, the IV regression includes dummy variables for deciles of firm size, formed at the end of March, interacted with year dummies. With these fixed effects, the instrument is leveraging the fact that firms which are similar in size in March, but have differential exposure to market returns from April to late May (based on their CAPM beta) will end up in different indices for index families that rebalance around the end of June (e.g., Russell and S&P). This alternative instrumentation approach is useful because it does not condition on future index membership and because it exploits a different source of variation than the two IVs above (cross-sectional vs. time series).

For this instrument, the exclusion restriction is that a firm’s CAPM beta times the market return from April to May is exogenous to price informativeness in year following July. This assumption would be less plausible if stocks with high beta had high idiosyncratic volatility. To partially address this concern, I explicitly control for idiosyncratic volatility over the

period used to compute CAPM beta. Reassuringly, in this alternative IV, the first stage and reduced form are both statistically significant, and the causal estimates are comparable in magnitude to those found in this section.

5 Mechanisms

My preferred explanation for why passive ownership decreases pre-earnings announcement price informativeness is that passive investors gather less firm-specific information. To support this claim, I start by showing that the negative relationship between passive ownership and pre-earnings announcement price informativeness is coming from the firm-specific component of information. Then, I present both direct and indirect evidence which suggests that passive owners demand less information about firm specific news. Next, I leverage pre-earnings announcement trading volume to distinguish between theories that relate private information gathering to trade in financial markets. Finally, I discuss why the equilibrium response of non-passive investors doesn't fully offset the effects of passive ownership.

5.1 Systematic vs. idiosyncratic news

Recent work on the effect of passive ownership on information gathering has highlighted a trade off that passive investors face in terms of the information they collect (Cong et al. (2020), Glosten et al. (2021)). Because passive investors have diversified portfolios, systematic news is more important to them, so they will optimally collect more systematic information and less firm-specific news. To test these theories, I examine the response of stock prices to news of a specific size. The intuition is that if investors have less precise beliefs before an announcement, they will update significantly afterwards, leading to a larger price change (Ganuza and Penalva, 2010). Given that passive investors are less likely to collect stock-specific information, this effect should be especially strong for the firm-specific component of news. This yields a testable prediction for the effect of passive ownership on earnings responses.

Prediction 4: *Stocks with more passive ownership should respond more to earnings news of a given size. This effect should be especially strong for firm-specific news*

To test prediction 4, I use the following earnings-response regression (Kothari and Sloan,

1992) to quantify the market’s reaction to a standardized measure of earnings news:

$$r_{i,t} = \alpha + \beta SUE_{i,t} + \theta Passive_{i,t} + \delta (SUE_{i,t} \times Passive_{i,t}) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \quad (8)$$

where $r_{i,t}$ denotes the market-adjusted return on the first day investors could trade on earnings information (in percentage points), Winsorized at the 1% and 99% level by year.²⁴ $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1, t-8)}(E_{i,t} - E_{i,t-4})}$ where $E_{i,t}$ is earnings-per-share from the IBES unadjusted detail file (i.e., “street” earnings). In words, 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.²⁵ In the Appendix, I show that the market reaction to earnings news of a given size is about 3 times as large now as it was in the early 1990s.

I also run versions of Equation 8 breaking SUE into positive and negative components and decomposing the earnings news into a systematic and idiosyncratic component using the method in Glosten et al. (2021). This is done by regressing firm-level 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.

Table 5 contains the regression results. Column 1 shows that, consistent with prediction 4, δ is positive and economically large, meaning that high passive ownership stocks are more responsive to earnings news. Column 2 shows that this effect is stronger for negative news than positive news. Column 3 shows that the increased responsiveness of high passive stocks to earnings news is concentrated in the firm-specific component, also consistent with prediction 4. Columns 4-6 confirm these results are robust to value weighting observations.

²⁴In the Appendix, I show that the results in this subsection are similar when instead using cumulative returns in windows of up to 5 days after the earnings-announcement.

²⁵I 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. Using this method, 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.

	(1)	(2)	(3)	(4)	(5)	(6)
<i>SUE</i>	2.64***			2.13***		
$\times Passive$	(0.24)			(0.26)		
$SUE \times 1_{SUE>0}$		1.15***			1.60***	
$\times Passive$		(0.37)			(0.42)	
$ SUE \times 1_{SUE \leq 0}$		-3.63***			-2.12***	
$\times Passive$		(0.34)			(0.63)	
$Sys.SUE \times 1_{Sys.SUE>0}$			1.40			1.00
$\times Passive$			(1.00)			(2.40)
$ Sys.SUE \times 1_{Sys.SUE \leq 0}$			0.67			1.5
$\times Passive$			(2.87)			(4.67)
$Idio.SUE \times 1_{Idio.SUE>0}$			1.04**			1.47***
$\times Passive$			(0.43)			(0.46)
$ Idio.SUE \times 1_{Idio.SUE \leq 0}$			-3.89***			-2.81***
$\times Passive$			(0.32)			(0.48)
Observations	333,814	333,814	333,814	333,814	333,814	333,814
R-squared	0.06	0.06	0.06	0.03	0.03	0.03
Firm + Year/Quarter FE	✓	✓	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓	✓	✓
Weight	Equal	Equal	Equal	Value	Value	Value

Table 5 Passive ownership and earnings responses. Estimates from:

$$r_{i,t} = \alpha + \beta SUE_{i,t} + \theta Passive_{i,t} + \delta (SUE_{i,t} \times Passive_{i,t}) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}$$

where $r_{i,t}$ is the market-adjusted return (in percentage points) on the effective earnings announcement date. Controls in $X_{i,t}$ include 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 and firm fixed effects. Standard errors double clustered at the firm and year-quarter level in parenthesis.

5.2 Information gathering

A natural explanation for a decrease in price informativeness is a decline in the share of informed investors or the precision of investors' signals (Grossman and Stiglitz (1980), Kyle (1985)). Passive managers, as well as investors in passive funds, lack strong incentives to gather and consume firm-specific information because these funds trade on mechanical rules, such as S&P 500 index membership (State Street's S&P 500 ETF Trust, SPY), or having 10 years of increasing dividend payments (Vanguard's Dividend Appreciation ETF, VIG). Further, because these funds are well diversified, even if they are traded by information-motivated investors, they are more likely to be used for bets on systematic, rather than firm-

specific, information. This logic yields an empirical prediction for the relationship between passive ownership and information gathering.

Prediction 5: *Passive ownership should cause firm-specific information gathering to decline*

One way to quantify information gathering is with Bloomberg terminal searches for specific tickers. As discussed by Ben-Rephael et al. (2017), these searches capture attention by institutional investors, who are the main users of Bloomberg’s products. The timing of when investors will search for information relative to earnings announcements, however, is not obvious. Attentive investors may search (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 channels, I perform a more general test. At the stock/month level, I ask whether stocks with more passive ownership have fewer Bloomberg terminal searches than stocks with less passive ownership. To this end, I run a regression of the continuous abnormal institutional attention measure from Ben-Rephael et al. (2017) (*AIAC*) on passive ownership.²⁶ The sample is stock/month observations between 2010 and 2018 that can be linked between Bloomberg and CRSP on ticker. All the controls and fixed effects are identical to Equation 3.

Columns 1 and 2 of Table 6 contains the results. Consistent with prediction 5, passive ownership is correlated with fewer Bloomberg searches. In terms of magnitudes, a 15% higher level of passive ownership implies -0.23 lower *AIAC*, which is about 24% of its whole sample mean of 0.97. If the mechanism behind this decline was just that passive investors gather no information, this estimate is roughly in line with the 15% decrease in information gathering we would expect ex-ante. Institutional investors (13F filers), which is what *AIAC* is designed to capture, however, only hold about 70% of the US stock market. So, if the rise of passive ownership was a re-allocation among institutional investors, we would expect to see a decline of $15\%/70\% \approx 21\%$, which is almost exactly what we see in Table 6.

As an alternative way to measure investors’ learning behavior, I examine downloads of SEC filings, with fewer downloads implying decreased gathering of fundamental information

²⁶These results are robust to instead using the other measures from Ben-Rephael et al. (2017) e.g., abnormal institutional attention (AIA) or the raw Bloomberg search intensity data.

	Bloomberg		Downloads		CAT	
	(1)	(2)	(3)	(4)	(5)	(6)
Passive Ownership	-1.52*** (0.37)	-2.29** (0.89)	-1.14*** (0.20)	-1.41** (0.65)	-11.49*** (3.21)	-10.25*** (2.77)
Observations	58,629	58,629	533,099	533,099	407,283	407,283
R-Squared	0.54	0.50	0.81	0.89	0.08	0.15
Firm + Year/Quarter FE	✓	✓	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value	Equal	Value

Table 6 Mechanisms. Estimates of β from:

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

For Columns 1-2, the left-hand side variable is *AIAC*, the measure of continuous abnormal institutional attention from Ben-Rephael et al. (2017) built on Bloomberg searches. For Columns 3-4 the left-hand side variable is one plus the natural logarithm of the number of non-robot downloads from Loughran and McDonald (2017). For Columns 5-6, the left-hand side is cumulative abnormal pre-earnings turnover. Controls in $X_{i,t}$ include 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 , and firm fixed effects ψ_i . Standard errors double clustered at the firm and year-quarter level in parenthesis.

(Loughran and McDonald, 2017). Specifically, I define $Downloads_{i,t}$ as one plus the natural logarithm of the number of non-robot downloads, measured using the method in Loughran and McDonald (2017) and obtained from their website. The sample runs from 2003-2015, excluding the data lost/damaged by the SEC from 9/2005-5/2006, and I match the downloads to CRSP/Compustat merged on CIK. As with the regressions using Bloomberg ticker searches, the unit of observation is firm-month.

Columns 3 and 4 of Table 6 are the analogue of Columns 1 and 2, but have $Downloads_{i,t}$ on the left-hand side. Also consistent with prediction 5, the estimated coefficient is negative and statistically significant, evidence that passive ownership is correlated with less investor attention.²⁷ In terms of magnitudes, a 15% higher level of passive ownership implies a decrease in $Downloads_{i,t}$ of -0.17, which is modest relative to its whole sample standard deviation of roughly 1.3. This magnitude, however, is harder to interpret than the results on

²⁷This result is consistent with Israeli et al. (2017) and Coles et al. (2022), who also show that passive ownership is negatively correlated with downloads of SEC filings.

Bloomberg searches, as we don't know who is downloading these SEC filings and whether or not they themselves are investors.

5.3 Sell-side analyst coverage

The relationship between passive ownership and information demand is intuitive. Investors buying an S&P 500 ETF probably care less about firm-specific fundamentals than people buying the underlying stocks. In addition to this direct effect on information gathering, however, a change in demand may have corresponding equilibrium effects on the supply of information, which can be measured using sell-side analyst coverage (Martineau and Zoican, 2021).

To fix ideas, suppose the results in the previous subsection imply that passive ownership shifts the demand curve for information inward. Given that information is not costless to produce, we expect it to have an upward sloping supply curve. All else equal, therefore, this decrease in demand should also lower the equilibrium supply of information.²⁸ This logic yields a testable prediction for the relationship between passive ownership and information supply, as measured by sell-side analyst coverage.

Prediction 6: *Passive ownership should be correlated with decreased quantity and quality of sell-side analyst coverage*

To test prediction 6, I run versions of my baseline OLS regression (Equation 3) with measures of information production by sell-side analysts on the left-hand side. 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, consistent with prediction 6. This mirrors Israeli et al. (2017)

²⁸It's possible that the supply curve for information also shifts in response to rising passive ownership. Isolating this effect is difficult, however, without a way to measure the price of information e.g., the cost of analyst reports. My results, therefore, can only speak to the net effect of passive ownership on the supply of and demand for information.

and Coles et al. (2022), who also show that ETF ownership is negatively correlated with the number of analyst estimates. Column 2 shows that passive ownership is correlated with a larger standard deviation of analyst estimates. Increased forecast dispersion is evidence of more uncertainty about the fundamental value of these firms (Diether et al. (2002), Zhang (2006)), which is also consistent with prediction 6.

One concern with these results, however, is that 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 inaccuracy which explicitly accounts for the increase in dispersion. Specifically, I define inaccuracy as 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 about high passive stocks, we would expect their forecasts to be less accurate, even when accounting for the increase in average uncertainty. Column 3 shows that this prediction holds empirically.

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 estimates 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 more passive ownership. Both Columns 5 and 6 imply analysts are expending less effort gathering information on stocks with more passive ownership, which also corroborates prediction 6.

The findings in Table 7 seem at odds with results in Section 4 because 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. Consistent with this hypothesis, in the Appendix I show that even though S&P 500 index addition leads to increased analyst coverage, it also leads to increased dispersion in analyst forecasts and decreased analyst accuracy. Moving from the Russell 1000 to the 2000 causes a drop in analyst coverage and accuracy, but this may be because these firms are shrinking in size.

	Num. Est (1)	SD(Est.) (2)	Dist./SD(Est.) (3)	Updates (4)	Time (5)
Passive Ownership	-11.64*** (1.38)	0.72*** (0.17)	1.97*** (0.45)	-0.45*** (0.10)	0.35*** (0.11)
Observations	216,805	216,805	216,805	133,176	133,176
R-squared	0.79	0.64	0.13	0.26	0.55
Mean	8.62	0.09	2.23	2.23	3.76
St. Dev.	5.94	0.41	2.97	0.45	0.84

Table 7 Passive ownership and coverage by sell-side analysts. Estimates of β from:

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

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, 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 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 and firm fixed effects. Standard errors double clustered at the firm and year-quarter level in parenthesis. The last two rows of the table present the means and standard deviations of the left-hand side variables.

5.4 Pre-earnings abnormal turnover

Next, I study the relationship between passive ownership and pre-earnings announcement trading volume. This is useful for distinguishing between theories that relate private information collection to trade in financial markets. While the empirical quantity of volume is difficult to rationalize (Cochrane, 2004), a mechanism common to many models is that information is a key motivation for trade. In some models, private information can increase trading volume as it is a source of heterogeneity among investors, and such disagreement makes them willing to trade (Wang, 1994). Too much private information can, however, decrease volume as fears of adverse selection deter uninformed investors from trading (Foster and Viswanathan (1990), Foster and Viswanathan (1993)).

One challenge with bringing these theories to the data is that private information is hard to quantify. My results show, however, that passive ownership decreases pre-earnings announcement information gathering. Therefore, the relationship between passive ownership

and pre-earnings announcement trading volume can speak to the relative strength of different channels proposed in the literature. This exercise is similar in spirit to Manela (2014), who examines trading and returns around a different set of public information release events (FDA drug approvals) to distinguish between the competing effects of the speed with which information diffuses through financial markets.

To quantify pre-earnings trading volume, 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:

$$AT_{i,t+\tau} = \frac{T_{i,t+\tau}}{\overline{T_{i,t-22}}} = \frac{T_{i,t+\tau}}{\sum_{k=1}^{252} T_{i,t-22-k} / 252} \quad (9)$$

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 amplifying or dampening changes in trading.

In the Appendix, I show that there has been a drop in trading volume throughout the month before earnings announcement over the past 3 decades. To summarize this decline, I define cumulative abnormal pre-earnings turnover as:

$$CAT_{i,t} = \sum_{\tau=-22}^{-1} AT_{i,t+\tau} \quad (10)$$

In words, CAT is the sum of abnormal turnover from $t - 22$ to $t - 1$ for firm i around earnings date t . To reduce the influence of outliers, I Winsorize CAT at the 1% and 99% level by year. Between the 1990s and 2010s, average $CAT_{i,t}$ declined by about 1, which 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.

I run a regression of CAT on passive ownership with the same controls and fixed effects

as Equation 3.²⁹ The results are in Columns 5 and 6 of Table 6, which show a strong negative relationship between passive ownership and pre-earnings announcement abnormal turnover. In terms of magnitudes, a firm in the 90th percentile of passive ownership in 2018 has a -2.6 lower *CAT* than a firm in the 10th percentile of passive ownership in 2018.

Returning to the theories discussed above, given that passive ownership decreases information gathering, we would expect it to decrease investor heterogeneity and therefore decrease trading volume. On the other hand, decreased information gathering should decrease adverse selection, which would tend to increase trading volume. The negative empirical relationship between passive ownership and pre-earnings trading volume suggests that in my setting, the effect of decreased disagreement tends to dominate the effect of decreased adverse selection.

One concern is that these regression results are mechanical functions of passive ownerships' effect on average trading volume. Passive investors may trade less and therefore more passive ownership leads to less trading overall. The way *CAT* is defined, however, should prevent passive ownership from causing a mechanical decrease in pre-earnings announcement trading volume, because passive ownership's effect of lowering average trading would be incorporated into past turnover (i.e., the denominator of Equation 9). 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.

5.5 Equilibrium response of non-passive investors

Suppose passive investors gather no stock-specific information. Then, as passive ownership increases, we would mechanically expect total information gathering to decrease, holding fixed the behavior of the remaining non-passive investors. If pre-earnings announcement prices have become less informative, however, the returns to becoming informed should have increased. So, a question remains as to why the remaining non-passive investors don't increase their information production to capitalize on this, as occurs in the model of Coles et al. (2022).

A natural reason why non-passive investors wouldn't fully compensate for the decline in

²⁹*CAT* is similar to Manela (2014)'s measure of cumulative abnormal turnover (*CATO*) before FDA drug approvals. In the Appendix I show I obtain similar results using *CATO* instead of *CAT* and I provide a more detailed explanation of *CAT*'s advantages in my setting.

information production is that passive ownership’s presence makes it harder to profit from private information. This might apply in my setting because as discussed in Ben-David et al. (2018), ETFs (but not non-ETF index funds) increase non-fundamental volatility in the underlying stocks. This could deter informed investors from gathering information, as there is some chance that before the end of their investment horizon, they are hit with a large volatility shock, which forces them to sell at a loss (De Long et al., 1990). An implication of this is that the effects I document in Section 3 should be stronger for ETFs than non-ETF passive funds.

To test this hypothesis, I re-run the baseline OLS regressions (Equation 3), but break passive ownership into ETFs and all passive funds that are not ETFs (i.e., index mutual funds). Panel A of Table 8 shows that ETFs have a larger effect on $DM_{i,t}$ and $QVS_{i,t}$ than non-ETF passive funds. These results are consistent with ETFs increasing the limits to arbitrage by boosting non-fundamental volatility, which lowers the equilibrium response of non-passive investors.

One concern with the results in Panel A of Table 8 is that, because the coefficient on non-ETF passive ownership is insignificant, ETFs explain all the effects of passive ownership on pre-earnings announcement price informativeness. Separating the effects of index mutual funds from ETFs is difficult, however, as they have a correlation coefficient of almost 0.7. Therefore, the coefficient on non-ETF passive ownership could be insignificant because of collinearity with ETF ownership. As an additional check, in Panel B of Table 8, I replicate Panel A, but only include non-ETF passive ownership on the right-hand side. This restores the statistical significance of passive mutual fund ownership, evidence which suggests that my results are not entirely driven by a feature specific to ETFs.

Another possible channel is that passive ownership decreases pre-earnings liquidity. Consistent with this, as shown in Table 6, stocks with more passive ownership have relatively less pre-earnings trading volume. One explanation for decreased liquidity is that the nature of passive ownership makes it harder to hide informed orders. In models like Kyle (1985), the market maker cannot tell whether demand is coming from insiders or noise traders. Unlike this, at the end of every day, investors can observe the exact change in the number of shares held by ETFs. So, if more volume is coming from ETFs, it might be harder to profit from private information in the pre-earnings period, as other investors will be able to detect someone trading on information and push prices against them.

Panel A: ETFs and Non-ETF Indexers				
	QVS		DM	
	(1)	(2)	(3)	(4)
ETF	-61.62*** (5.13)	-29.06** (13.97)	-6.68*** (0.86)	-6.00*** (1.63)
Non-ETF Passive	1.40 (5.79)	-17.47 (20.48)	-1.21 (1.25)	-2.10 (2.50)
Observations	429,672	429,672	429,672	429,672
R-Squared	0.232	0.244	0.222	0.276
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value
Panel B: Only Non-ETF Indexers				
	QVS		DM	
	(1)	(2)	(3)	(4)
Non-ETF Passive	-36.37*** (6.59)	-45.66** (20.84)	-5.30*** (1.33)	-7.91*** (2.70)
Observations	429,672	429,672	429,672	429,672
R-Squared	0.231	0.244	0.221	0.276
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value

Table 8 Breakdown of QVS and DM regression results by type of passive ownership.
Table with estimates of b_i s from:

$$PriceInformativeness_{i,t} = \alpha + b_1 ETF_{i,t} + b_2 Non-ETF\ Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$$

where $PriceInformativeness_{i,t}$ is either DM or QVS , $ETF_{i,t}$ is ETF ownership and $Non-ETF\ Passive_{i,t}$ is ownership by non-ETF passive funds. Controls in $X_{i,t}$ include 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 ownership measures are expressed as decimals, so 0.01 = 1% of firm i 's shares are owned by ETFs. Standard errors double clustered at the firm and year-quarter level in parenthesis.

Risk and liquidity are just two of many possible explanations for why non-passive investors don't fully compensate for the lack of information gathering by passive investors.³⁰ These general equilibrium effects, however, are hard to measure. The evidence in Table 6 only speaks to net changes in information demand and, more broadly, the regressions in Tables 2 and 4 only speak to the net effect of passive ownership on pre-earnings announcement price informativeness. So, it is possible that non-passive investors respond by gathering more information, just not enough to fully offset the decrease coming from passive investors. Without being able to see individual investors' attention, however, it is difficult to quantify such effects.

6 Conclusion

In this paper, I propose two ways to measure the fraction of earnings information incorporated into prices before the announcement itself. I show that over the past 30-years, pre-earnings announcement price informativeness has been steadily declining. Passive ownership played an important role in this trend, as taking the average of the point estimates from the OLS and both IVs implies that a 15% increase in passive ownership decreases QVS by 14.87, a roughly 16% decline relative to its 1990 mean.

My proposed mechanism is that passive investors gather less firm-specific information. To support this claim, I first use earnings response regressions to show that the decline in pre-earnings announcement price informativeness came from firm-specific news. Then, I show direct evidence of decreased information gathering for high passive stocks through Bloomberg terminal searches, downloads of SEC filings and sell-side analyst coverage. Finally, I argue why passive ownership may increase the limits to arbitrage, which prevents non-passive investors from fully offsetting these effects.

Relative to total institutional ownership, passive ownership is still small, owning only about 15% of the US stock market. Even at this level, passive ownership has led to significant

³⁰Analogously, Haddad et al. (2021) find empirically that non-passive investors do not fully offset passive ownership's tendency to decrease demand elasticity. They propose several explanations for this, including costly information acquisition, risk, institutional mandates, bounded rationality and strategic interactions among investors (e.g., price impact and herding). In addition, Bond and Garcia (2018) develop a model where non-passive investors do not fully offset passive ownership's effect on price informativeness owing to participation costs and strategic complementary in participation decisions.

changes in how stock prices anticipate the information contained in earnings announcements. As passive ownership continues to grow, these effects may be amplified, further changing the way equity markets reflect firm-specific information.

References

- Aghion, P., Van Reenen, J., and Zingales, L. (2013). Innovation and institutional ownership. *American economic review*, 103(1):277–304.
- Antoniou, C., Li, F. W., Liu, X., Subrahmanyam, A., and Sun, C. (2020). The real effects of exchange-traded funds. *Available at SSRN 3129369*.
- Appel, I., Gormley, T. A., and Keim, D. B. (2020). Identification using russell 1000/2000 index assignments: A discussion of methodologies. *Critical Finance Review, Forthcoming*.
- Appel, I. R., Gormley, T. A., and Keim, D. B. (2016). Passive investors, not passive owners. *Journal of Financial Economics*, 121(1).
- Asquith, P., Pathak, P. A., and Ritter, J. R. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics*, 78(2):243–276.
- Åstebro, T. and Penalva, J. (2022). Two-sided platform governance: Are founders manipulating the crowd in crowdfunding? *Available at SSRN 3991200*.
- Azar, J., Schmalz, M. C., and Tecu, I. (2018). Anticompetitive Effects of Common Ownership. *Journal of Finance*, 73(4).
- Bai, J., Philippon, T., and Savov, A. (2016). Have financial markets become more informative? *Journal of Financial Economics*, 122(3).
- Ball, R. and Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, 6(2).
- Bebchuk, L., Cohen, A., and Ferrell, A. (2009). What matters in corporate governance? *The Review of financial studies*, 22(2):783–827.
- Beber, A. and Brandt, M. W. (2006). The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, 53(8):1997–2039.
- 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*.
- Ben-Rephael, A., Da, Z., and Israelsen, R. D. (2017). It depends on where you search: Institutional investor attention and underreaction to news. *The Review of Financial Studies*, 30(9):3009–3047.
- Bennett, B., Stulz, R., and Wang, Z. (2020a). Does the stock market make firms more productive? *Journal of Financial Economics*, 136(2):281–306.
- Bennett, B., Stulz, R. M., and Wang, Z. (2020b). Does joining the s&p 500 index hurt firms? Technical report, National Bureau of Economic Research.
- Bernstein, S. (2015). Does going public affect innovation? *The Journal of finance*, 70(4):1365–1403.
- 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*.
- Bond, P. and Garcia, D. (2018). The equilibrium consequences of indexing. *Review of Financial Studies*, forthcoming.
- Boone, A. L. and White, J. T. (2015). The effect of institutional ownership on firm transparency and information production. *Journal of Financial Economics*, 117(3).
- Breugem, M. and Buss, A. (2019). Institutional investors and information acquisition: Implications for asset prices and informational efficiency. *Review of Financial Studies*, 32(6).
- Brogaard, J., Ringgenberg, M. C., and Sovich, D. (2019). The economic impact of index investing. *The Review of Financial Studies*, 32(9):3461–3499.
- 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).
- Campbell, J. Y. and Shiller, R. J. (1988). Stock prices, earnings, and expected dividends. *the Journal of Finance*, 43(3):661–676.
- 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).
- Chernozhukov, V. and Hansen, C. (2008a). Instrumental variable quantile regression: A robust inference approach. *Journal of Econometrics*, 142(1):379–398.
- Chernozhukov, V. and Hansen, C. (2008b). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1):68–71.
- Chung, K. H. and Zhang, H. (2011). Corporate governance and institutional ownership. *Journal of financial and quantitative analysis*, 46(1):247–273.
- Cochrane, J. H. (2004). Asset pricing: Liquidity, trading, and asset prices. *NBER Reporter Online*, (Winter 2004/05):1–12.
- Coffee, J. C. (2007). Law and the market: The impact of enforcement.
- Cohen, L. and Lou, D. (2012). Complicated firms. *Journal of financial economics*, 104(2):383–400.
- Cohen, R. B., Polk, C., and Vuolteenaho, T. (2003). The value spread. *The Journal of Finance*, 58(2):609–641.
- Coles, J. L., Heath, D., and Ringgenberg, M. C. (2022). On index investing. *Journal of Financial Economics*, 145(3):665–683.
- Cong, L. W., Huang, S., and Xu, D. (2020). Rise of factor investing: security design and asset pricing implications. Technical report, Working Paper.
- Dannhauser, C. D. (2017). The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). *Journal of Financial Economics*, 125(3).

- Dávila, E. and Parlatore, C. (2018). Identifying Price Informativeness. *National Bureau of Economic Research Working Paper Series*.
- Dávila, E. and Parlatore, C. (2021). Trading Costs and Informational Efficiency. *Journal of Finance*, 76(3).
- De Long, J. B., Shleifer, A., Summers, L. H., and Waldmann, R. J. (1990). Noise trader risk in financial markets. *Journal of political Economy*, 98(4):703–738.
- DeLisle, R. J., French, D. W., and Schutte, M. G. (2017). Passive institutional ownership, r2 trends, and price informativeness. *Financial Review*, 52(4):627–659.
- 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.
- Ernst, T. (2020). Stock-Specific Price Discovery From ETFs. *Working Paper*, 1(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, F. D. and Viswanathan, S. (1990). A theory of the interday variations in volume, variance, and trading costs in securities markets. *The Review of Financial Studies*, 3(4):593–624.
- Foster, F. D. and Viswanathan, S. (1993). Variations in trading volume, return volatility, and trading costs: Evidence on recent price formation models. *The Journal of Finance*, 48(1):187–211.
- Ganuzza, J. J. and Penalva, J. S. (2010). Signal orderings based on dispersion and the supply of private information in auctions. *Econometrica*, 78(3):1007–1030.
- 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*.
- Gloßner, S. (2019). Investor horizons, long-term blockholders, and corporate social responsibility. *Journal of Banking and Finance*, 103.
- Glosten, L., Nallareddy, S., and Zou, Y. (2021). ETF activity and informational efficiency of underlying securities. *Management Science*, 67(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*.
- Hanson, S. G. and Sunderam, A. (2014). The growth and limits of arbitrage: Evidence from short interest. *The Review of Financial Studies*, 27(4):1238–1286.
- Heath, D., Ringgenberg, M. C., Samadi, M., and Werner, I. M. (2020). Reusing natural experiments. *Available at SSRN 3991200*.
- 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., Sundaresan, S., and Wang, T. (2018a). Do foreign investors improve market efficiency? Technical report, National Bureau of Economic Research.
- Kacperczyk, M. T., Nosal, J. B., and Sundaresan, S. (2018b). 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*.
- Lochner, L. and Moretti, E. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *American economic review*, 94(1):155–189.
- Loughran, T. and McDonald, B. (2017). The use of edgar filings by investors. *Journal of Behavioral Finance*, 18(2):231–248.
- Madhavan, A. (2003). The russell reconstitution effect. *Financial Analysts Journal*, 59(4):51–64.
- Madhavan, A. (2014). Exchange-traded funds: an overview of institutions, trading, and impacts. *Annu. Rev. Financ. Econ.*, 6(1):311–341.
- Madhavan, A. N. (2016). *Exchange-traded funds and the new dynamics of investing*. Oxford University Press.
- Malikov, G. (2020). Information, Participation, and Passive Investing. *SSRN Electronic Journal*.
- Manela, A. (2014). The value of diffusing information. *Journal of Financial Economics*, 111(1):181–199.
- Martineau, C. and Zoican, M. (2021). Investor attention and the cross-section of analyst coverage. *Available at SSRN 3376162*.
- Massa, M., Schumacher, D., and Wang, Y. (2021). Who Is Afraid of BlackRock? *The Review of Financial Studies*, 34(4).
- Mauboussin, M. (2019). Looking for Easy Games in Bonds. *Bluemountain Investment Research*.
- 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*.

- O'Brien, P. C. and Bhushan, R. (1990). Analyst following and institutional ownership. *Journal of Accounting Research*, 28:55–76.
- Pavlova, A. and Sikorskaya, T. (2022). Benchmarking intensity.
- Qin, N. and Singal, V. (2015). Indexing and stock price efficiency. *Financial management*, 44(4):875–904.
- Roll, R. (1988). R-squared. *Journal of finance*, 43(2):541–566.
- Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.
- Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.
- Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression.
- Velury, U. and Jenkins, D. S. (2006). Institutional ownership and the quality of earnings. *Journal of business research*, 59(9):1043–1051.
- Wang, J. (1994). A Model of Competitive Stock Trading Volume. *Journal of Political Economy*, 102(1).
- Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *Review of Financial Studies*, 31(6).
- Wurgler, J. (2010). On the economic consequences of index-linked investing. Technical report, National Bureau of Economic Research.
- Zhang, X. F. (2006). Information uncertainty and stock returns. *The Journal of Finance*, 61(1):105–137.
- Zou, Y. (2018). Lost in the rising tide: Etf flows and valuation.

Internet Appendix

A Stylized Facts

A.1 Visualizing the pre-earnings announcement drift

Ball and Brown (1968) show that prices tend to drift up before the release of good news and drift down before the release of bad news. Visualizing this requires a definition of good and bad news, so following Novy-Marx (2015), define standardized unexpected earnings (SUE) as:

$$SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})} \quad (11)$$

where $E_{i,t}$ denotes earnings per share for firm i in quarter t in the IBES Unadjusted Detail File. In words, Equation 11 is measuring the year-over-year (YOY) change in earnings, divided by the standard deviation of YOY changes in earnings over the past 8 quarters. Each quarter, I sort firms into deciles of SUE and calculate the cumulative market-adjusted returns of a \$1 investment 22 trading days before the earnings announcement.

Figure 3 plots these average cumulative market-adjusted returns by SUE decile for two different time periods: 2001-2007 and 2010-2018. 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. Consistent with Ball and Brown (1968), in the left panel, the best news is preceded by positive market-adjusted returns while the worst news is preceded by negative market-adjusted returns.

Consistent with the trend in Figure 1 in the main body of the paper, firms in each decile move less before earnings days between 2010 and 2018 than between 2001 and 2007. 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.

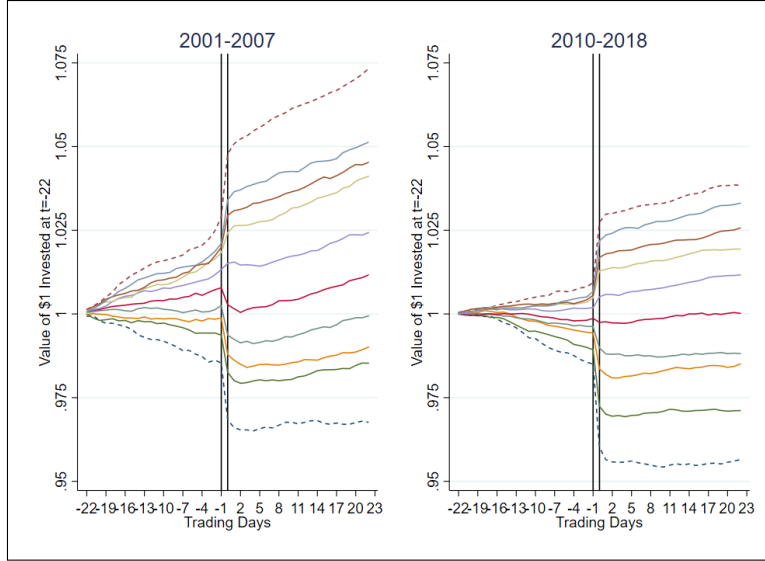


Figure 3. Decline of pre-earnings drift by SUE decile. Each quarter, firms are sorted into deciles based on standardized unexpected earnings (*SUE*). Each line represents the cross-sectional average market-adjusted return of \$1 invested at $t = -22$. The brown dashed line represents the average for 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.

A.2 Decomposition of earnings days' share of volatility

Figure 4 decomposes the decline of *QVS* into the rise in volatility on earnings days and the decline in volatility on all other days (which determines the sum of squared returns from $t - 22$ to t i.e., the denominator of *QVS*). The trend in *QVS* was driven by a simultaneous increase in earnings day volatility (left panel) and a decrease in volatility on all other days (right panel).

A.3 Relationship between *DM* and *QVS*

Given the similar time-series trends in *QVS* and *DM*, a natural question is whether they capture different information. By construction, they will both tend to be lower if the earnings-day return is large in absolute value. But, as discussed in the main body, they may sometimes yield different conclusions about pre-earnings announcement price informativeness because e.g., *DM* is sensitive to the level of volatility, while *QVS* is not. In terms of their

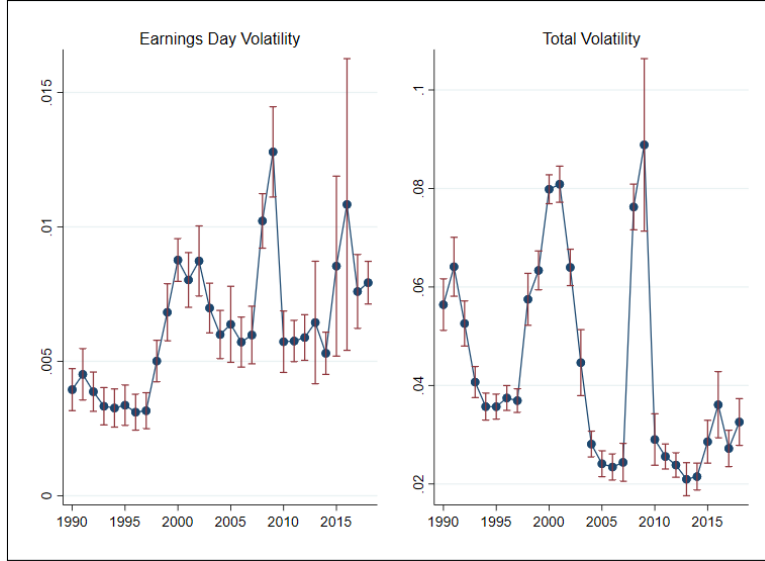


Figure 4. Decomposition of QVS . This figure plots coefficients from a regression of the pieces of QVS on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient. The left panel has the squared earnings-day return on the left-hand side while the right panel has the sum of squared returns from $t - 22$ to t on the left-hand side. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.

statistical relationship, a univariate regression of QVS on DM has an R-squared of just under 50%. To visualize this relationship, Figure 5 presents a scatter plot with QVS on the y-axis and DM on the x-axis.

B Data details

B.1 Details on construction of control variables

One month lagged market capitalization: Market capitalization of the stock at the end of the calendar month before the month of the earnings announcement

Time since listing: Time (in years) since security first appeared in CRSP

Returns from month $t - 12$ to $t - 2$: Cumulative geometric returns from month $t - 12$ to $t - 2$, where t is the month of the earnings announcement. This is flagged as missing if a firm has more than 4 observations with missing returns over the $t - 12$ to $t - 2$ period.

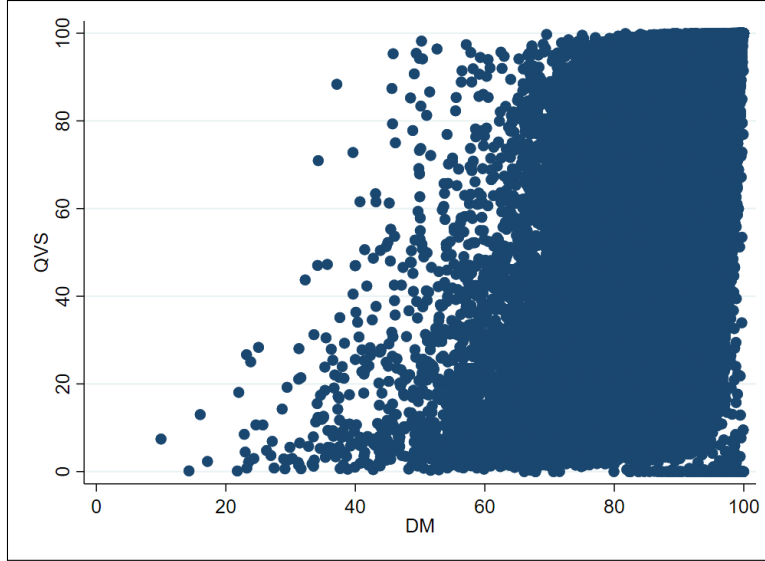


Figure 5. QVS vs. DM. This figure is a scatter plot of QVS on DM . Each blue dot represents a single earnings announcement.

Lagged book-to-market ratio: Book to market ratio of the stock at the end of the calendar month before the month of the earnings announcement from the WRDS financial ratios suite.
Total institutional ownership: The fraction of a stock's shares outstanding held by all 13-F filing institutions. Computed using the code [here](#).

CAPM beta, total volatility (sum of squared returns), idiosyncratic volatility (sum of squared CAPM residuals) and CAPM R-squared are all from the WRDS beta suite and are computed over the previous 252 trading days. For a firm to be included, it must have at least 151 non-missing returns over this period.

B.2 IBES

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

IBES release times is likely not driving my OLS estimates because in the main body of the paper, when ruling out the influence of Regulation Fair Disclosure, I show my results are quantitatively unchanged using only post-2000 data (i.e., the subsample where there are few missing earnings release times in IBES).

B.3 Computing passive and institutional ownership

To calculate passive ownership, I need to identify the holdings of passive funds, which I obtain from the Thompson S12 data.³¹ I use the WRDS MF LINKS database to connect the funds identified as passive in CRSP with the S12 data. If a security never appears in the S12 data, I assume its passive ownership is zero unless the firm is also considered to have missing institutional ownership by this code (IO_MISSING = 1), in which case I also set passive ownership to missing. S12 data is only reported at the end of each calendar quarter, so to get a monthly estimate of passive ownership, I linearly interpolate passive ownership between quarter-ends. All results are quantitatively unchanged if I instead fix passive ownership at its last reported level between the ends of calendar quarters.

C Cross-Sectional Regressions

C.1 Non-linear relationship between size and pre-earnings announcement price informativeness

Although all the baseline regressions explicitly control for market capitalization, one might be worried that firm size has a non-linear effect on pre-earnings announcement price informativeness. To ameliorate this concern, Table 9 replicates the baseline results, removing the control for market capitalization and instead including dummy variables for deciles of market capitalization, formed at the end of the previous calendar quarter. The results are quantitatively unchanged by including these fixed effects, suggesting that a non-linear effect of size on pre-earnings announcement price informativeness is not driving my OLS results.

³¹The S12 database is constructed from a combination of mutual funds' voluntary reporting and SEC filings on which securities they hold.

	QVS		DM	
	(1)	(2)	(3)	(4)
Passive Ownership	-42.06*** (3.03)	-33.52*** (10.06)	-5.15*** (0.62)	-5.73*** (1.31)
Observations	430,489	430,489	430,489	430,489
R-Squared	0.23	0.24	0.22	0.28
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value

Table 9 Cross-sectional regression of price informativeness on passive ownership (fixed effects for deciles of market capitalization). Table with estimates of β from:

$$\text{Price informativeness}_{i,t} = \alpha + \beta \text{Passive}_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$$

where Price informativeness $_{i,t}$ is either $QVS_{i,t}$ or $DM_{i,t}$. Controls in $X_{i,t}$ include age, 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 specification also include dummy variables for deciles of market capitalization, formed each quarter. 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.

C.2 Stocks with zero passive ownership

Previous studies on passive investing have excluded securities with zero passive ownership (Dannhauser, 2017). The logic is that there may be something special about the subset of securities passive funds avoid holding. Table 10 contains the summary statistics and Table 11 contains the baseline OLS regression results for the subsample with strictly positive passive ownership. Neither table is quantitatively changed from the corresponding one in the main body of the text by excluding these observations.

This filter does not have a large effect on my results because it shrinks my total sample size by less than 10%. This is because Vanguard's Total Stock Market ETF was launched in 2001, and this index is designed to track the CRSP US Total Market Index, which includes almost all ordinary common shares traded on major exchanges. I am already restricting to this universe of stocks (in addition to requiring a match between CRSP and IBES), so there are few stocks in this set that have zero passive ownership after 2001.

		25%	50%	Mean	75%	St. Dev.
<i>QVS</i>	1990-1999	89.51	96.85	91.20	99.41	14.02
<i>DM</i>		95.58	97.90	96.58	99.16	4.22
<i>Passive</i>		0.19	0.53	0.82	1.21	0.84
<i>QVS</i>	2010-2018	59.95	87.69	76.13	97.77	26.38
<i>DM</i>		93.33	96.76	95.06	98.67	5.42
<i>Passive</i>		4.11	8.76	9.25	13.13	6.38
<i>QVS</i>	All Years	79.34	94.50	84.57	98.98	21.49
<i>DM</i>		94.11	97.26	95.52	98.90	5.29
<i>Passive</i>		0.55	2.06	4.06	5.79	4.99

Table 10 Summary Statistics (dropping observations with zero passive ownership). Cross-sectional equal-weighted means, standard deviations and distributions of price informativeness and passive ownership. Excludes all observations with zero passive ownership.

	QVS		DM	
	(1)	(2)	(3)	(4)
Passive Ownership	-37.88*** (3.20)	-20.61** (10.05)	-4.50*** (0.63)	-5.15*** (1.40)
Observations	401,733	401,733	401,733	401,733
R-Squared	0.23	0.24	0.23	0.28
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value

Table 11 Cross-sectional regression of price informativeness on passive ownership (Dropping observations with zero passive ownership). Table with estimates of β from:

$$\text{Price informativeness}_{i,t} = \alpha + \beta \text{Passive}_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$$

where Price informativeness_{*i,t*} is either *QVS*_{*i,t*} or *DM*_{*i,t*}. Controls in *X*_{*i,t*} include 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. Excludes all observations with zero passive ownership.

C.3 Comparison to previous work

While not identical, DM is similar to the price-jump measure of Weller (2018), which is also designed to capture the fraction of earnings information incorporated into prices before it was formally released. The difference is that DM uses gross returns, while price-jump uses net returns. To fix ideas, consider a net return version of DM : $\widehat{DM}_{i,t} = 100 \times r_{t-22,t-1} / r_{t-22,t}$. $\widehat{DM}_{i,t}$ solves one issue with $DM_{i,t}$ in that it is symmetric with respect to positive and negative returns.

$\widehat{DM}_{i,t}$, however, has two drawbacks. The first is that, like $DM_{i,t}$, $\widehat{DM}_{i,t}$ is sensitive to the level of volatility. Further, the mean of $\widehat{DM}_{i,t}$ may not be well defined, as $r_{t-22,t}$ can be equal to zero. Weller (2018) overcomes this challenge by filtering out “non-events”, defined as observations with $r_{t-22,t}$ close to zero, which constitute almost 50% of earnings announcements in his sample. This filter, however, can complicate any analysis where the right-hand side variable of interest is related to market capitalization, as is the case with passive ownership, or when using value weights, because non-events are not evenly spread across the firm size distribution.

DM and QVS are also related to absolute $CARs$ around earnings announcements (Ball and Brown, 1968) and pre- and post- drug approval $CARs$ (Manela, 2014). I believe that in my setting, as discussed in Weller (2018), DM and QVS have the advantage that they captures the share of information incorporated into prices before it is formally announced. In the next two sections, I show that using Weller’s price-jump measure or $CARs$ does not change any of my empirical conclusions.

C.3.1 Relation to Weller (2018)

Weller (2018) studies the effect of algorithmic trading (AT) activity on information gathering. The logic is that algorithmic traders can reduce the returns to gathering information by back-running informed investors. If this deters information acquisition, we would expect stocks with more AT activity to have less informative prices. He quantifies pre-earnings announcement price informativeness using the price jump, defined as:

$$jump_{i,t}^{(a,b)} = \frac{CAR_{i,t}^{(T-1,T+b)}}{CAR_{i,t}^{(T-a,T+b)}}$$

where $CAR_{i,t}^{(l,m)}$ is a cumulative abnormal return from day l to day m , $a = 21$ and $b = 2$. In words the price jump is fraction of the cumulative abnormal return from a days before the earnings announcement to b days after that occurs after the announcement itself. Although this is not identical to *DM* or *QVS*, the price jump is also designed to capture (one minus) the fraction of earnings information incorporated into prices before it is formally released. If less information is incorporated into prices ahead of time, we expect to observe large values of the price jump.

One limitation of $jump_{i,t}^{(a,b)}$ is that it is not defined when the company has near zero returns over the month leading up to and including the earnings announcement. Weller handles this issue by dropping earnings announcements where $CAR_{i,t}^{(T-a,T+b)}$ is small, which he calls the non-event filter. This filter, however, removes the majority of earnings announcements in his sample (54.5%). Because *DM* uses gross returns and *QVS* uses the sum of squared returns, they have the advantage that they can be computed for every earnings announcement.

C.3.2 Relation to Manela (2014)

Manela (2014) examines the relationship between the value of information and how fast that information diffuses through financial markets around a different set of news events: FDA drug approvals. One of the quantities he uses to study this relationship is pre vs. day-of vs. post announcement cumulative abnormal returns (CARs). The logic is that investors should trade more aggressively on faster-diffusing news. In equilibrium, this leads fast-diffusing news to have higher pre-announcement returns because of aggressive trading by informed insiders. On the other hand, slower diffusing news is mostly traded into prices after the announcement itself.

Although my setting is different, I can use the same logic to test whether passive ownership leads less information to be incorporated into prices before earnings announcements. By looking at drug approvals, Manela (2014) is focused on good news, so to create an analogue of these results for earnings announcements, I need to condition on the news itself. To this end, I split firms into 10 deciles based on their standardized unexpected earnings, and focus

on firms in the top decile.³² I then calculate the pre-announcement ($t = -5$ to $t = -1$), announcement-day ($t = 0$ to $t = 1$) and post-announcement ($t = 2$ to $t = 6$) cumulative market-adjusted returns (R_1 , R_2 and R_3). If passive ownership decreases the amount of good news incorporated into prices ahead of time, more will be incorporated into prices on the day of, increasing R_2 . An advantage of my measures relative to these CARs is that they ease comparison across stocks and time because *DM* and *QVS* capture the *share* of information incorporated into prices before it is formally announced.

C.3.3 Baseline OLS regressions with Weller (2018)’s and Manela (2014)’s measures of price informativeness

To test whether my results are sensitive to the way I defined pre-earnings announcement price informativeness, I re-run my baseline OLS regression with Weller (2018)’s and Manela (2014)’s measures on the left-hand side:

$$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t} \quad (12)$$

where $Outcome_{i,t}$ is either $jump_{i,t}^{(a,b)}$ (with $b = 2$ and $a = 22$), R_1 , R_2 or R_3 . Controls in $X_{i,t}$ include time since listing (age), one-month lagged market capitalization, returns from month $t - 12$ to $t - 2$, one-month lagged book-to-market ratio and the institutional ownership ratio. $X_{i,t}$ also includes CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility, all computed over the previous 252 trading days. Finally, the regression includes firm and year-quarter fixed effects. Standard errors are double-clustered at the firm and year-quarter level.

Column 1 of Table 12 shows that higher levels of passive ownership are correlated with larger price jumps. This implies that passive ownership leads to less informative pre-earnings announcement prices, consistent with my results using *DM* and *QVS*. Columns 2 and 4 show that high passive stocks that experience good news don’t tend to have larger pre-earnings or post-earnings returns. Column 3 shows, however, that they have larger average earnings-day returns. This suggests that high passive stocks had less of the good news incorporated into

³² $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t} - E_{i,t-4})}$ where $E_{i,t}$ denotes earnings per share for firm i in quarter t in the IBES Unadjusted Detail File. In words, *SUE* is measuring the year-over-year (YOY) change in earnings, divided by the standard deviation of YOY changes in earnings over the past 8 quarters.

Paper: Measure:	Weller (2018) Price Jump (1)	Manela (2014) R_1 R_2 R_3 (2) (3) (4)		
Passive Ownership	0.311*** (0.059)	0.0326 (0.021)	0.0706** (0.031)	-0.0359 (0.022)
Observations	148,931	31,571	31,571	31,571
R-Squared	0.183	0.262	0.244	0.243
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Equal	Equal	Equal

Table 12 Passive ownership and alternative measures of pre-earnings announcement price informativeness. Estimates of β from:

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

$Outcome_{i,t}$ is either $jump_{i,t}^{(22,2)}$, the price jump measure from Weller (2018) with $a = 22$ and $b = 2$, $R_{1,i,t}$, the cumulative market-adjusted return from $t = -5$ to $t = -1$, $R_{2,i,t}$, the cumulative market-adjusted return from $t = 0$ to $t = 1$ or $R_{3,i,t}$, the cumulative market-adjusted return from $t = 2$ to $t = 6$ from Manela (2014). Controls in $X_{i,t}$ include 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. In Columns 2-4, the sample is restricted to firms in the top decile of SUE.

their prices ahead of time, also consistent with the *DM* and *QVS* results in the main body of the paper.

C.4 Passive ownership's asymmetric effect on pre-earnings drift for positive vs. negative news

Figure 3 suggests that the time-trend toward decreased *DM* was not equal for firms which ended up releasing good news and firms which ended up releasing bad news. To clarify this asymmetry, Figure 6 presents a version of Figure 3 which splits stocks into quintiles of *SUE* and quartiles of passive ownership using data between 2010 and 2018. Figure 6 highlights two types of asymmetry. The first is that firms with low *SUE* have smaller pre-announcement

drift than firms with high SUE . The second is that the effect of passive ownership on the pre-earnings drift is stronger for firms that end up releasing bad news than good news.

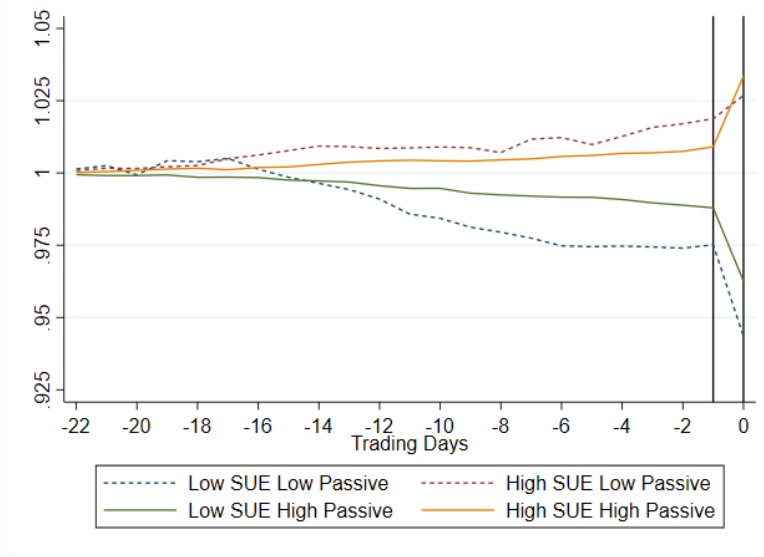


Figure 6. Pre-earnings drift by SUE quintile and quartile of passive ownership. Each quarter, firms are sorted into quintiles based on standardized unexpected earnings (SUE) and quartiles based on their passive ownership share. Each line represents the cross-sectional average market-adjusted return of \$1 invested at $t = -22$. Sample is earnings announcements between 2010-2018.

To quantify both of these effects, I run two regressions. The first is:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^5 b_j 1_{SUE_{i,t} \in Q_j} + \phi_t + \psi_i + e_{i,t} \quad (13)$$

where $1_{SUE_{i,t} \in Q_j}$ is an indicator variable equal to 1 if $SUE_{i,t}$ is in the j^{th} quintile of SUE in a given quarter (all results are similar running a version of Equation 13 with deciles of SUE instead of quintiles). Table 13 contains the results. Column 1 shows that, consistent with Figure 6, there is an unconditional asymmetry in DM between stocks which release good news and bad news. The last row of this column is the p-value from a test of $b_1 = b_5$, which suggests this difference is statistically significant. Column 3 shows the asymmetry is qualitatively unchanged by using value weights, instead of equal weights.

To quantify the role of passive ownership in this asymmetry (i.e., test for asymmetry

within a given level of passive ownership), I run a second regression which includes interaction terms between the quintiles of SUE and passive ownership:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^5 d_j 1_{SUE_{i,t} \in Q_j} + \sum_{j=1}^5 c_j 1_{SUE_{i,t} \in Q_j} \times Passive_{i,t} + \phi_t + \psi_i + e_{i,t} \quad (14)$$

Column 2 of Table 13 shows that there is an asymmetry between firms that release good and bad news for a given level of passive ownership, as c_1 is less than c_5 . The last row of this Column is the p-value from a test of $c_1 = c_5$, which again suggests the difference is statistically significant. Column 4 shows this is also robust to using value weights instead of equal weights.

One explanation for both asymmetries is shorting constraints. The logic is that an investor's hurdle rate for shorting may be higher than their hurdle rate for long-only investments. This is because there are frictions associated with short selling which are not present when buying a stock (e.g., the possibility of a short squeeze and getting margin called). So, when prices are too high, correcting them is harder than when prices are too low (Stambaugh et al. (2012) Stambaugh et al. (2015)).

If passive ownership makes prices less informative, we might expect that this effect would be stronger for firms which eventually release bad news. As I in the mechanisms section of the paper, there should be an equilibrium response of non-passive investors to the lack of information gathering by passive owners. Given shorting constraints, however, we might expect the equilibrium response to be larger for positive news than negative news, leading to an asymmetry in the pre-earnings drift within a given level of passive ownership.

These shorting frictions might be especially salient for stocks with more passive ownership because of the additional noise trader risk created by ETF arbitrage (as discussed in the mechanisms section of the paper and Ben-David et al. (2018)). Specifically, the remaining informed investors might be hesitant to short high passive stocks because prices are more likely to move against them in the meantime and have their short called. This may be true even if passive ownership increased the amount of shares available for shorting (Beschwitz et al., 2020), as the higher hurdle rates for shorts are mostly about short squeezes and margin calls rather than borrowing costs and share availability (Hanson and Sunderam, 2014).

	(1)	(2)	(3)	(4)
Passive	-3.600*** (0.743)	-3.260*** (0.760)	-5.431*** (1.546)	-5.072*** (1.696)
Low SUE	-0.457*** (0.042)	-0.291*** (0.046)	-0.124** (0.052)	-0.0295 (0.066)
2	-0.113*** (0.027)	-0.0829** (0.035)	-0.125** (0.061)	-0.0469 (0.075)
4	-0.123*** (0.028)	-0.185*** (0.040)	0.102** (0.050)	0.0353 (0.066)
High SUE	-0.318*** (0.033)	-0.410*** (0.046)	0.148*** (0.051)	0.118* (0.067)
Low SUE x Passive		-3.466*** (0.665)		-2.216** (1.093)
2 x Passive		-0.652 (0.478)		-1.701 (1.050)
4 x Passive		1.265** (0.523)		1.342 (0.874)
High SUE x Passive		1.893*** (0.602)		0.572 (0.919)
Observations	333,340	333,340	333,340	333,340
R-squared	0.216	0.216	0.254	0.254
p-Value	0.0040	0.0000	0.0000	0.0070
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓
Weight	Equal	Equal	Value	Value

Table 13 Effect of passive ownership on the pre-earnings drift by quintile of SUE.

Estimates of β , b_j , c_j and d_j from:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^5 b_j 1_{SUE_{i,t} \in Q_j} + \phi_t + \psi_i + e_{i,t}$$

where $1_{SUE_{i,t} \in Q_j}$ is an indicator variable equal to 1 if $SUE_{i,t}$ is in the j^{th} quintile of SUE in a given quarter.

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^5 c_j 1_{SUE_{i,t} \in Q_j} + \sum_{j=1}^5 d_j 1_{SUE_{i,t} \in Q_j} \times Passive_{i,t} + \phi_t + \psi_i + e_{i,t}$$

For every regression, the middle quintile is the omitted group. In columns 1 and 3, the last row of the table contains the p-value from a test of whether $b_1 = b_5$. In columns 2 and 4, the last row of the table contains the p-value from a test of whether $c_1 = c_5$. 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.

C.5 Robustness to including a longer post-earnings announcement window

One concern is that my results are specific to only including the effective earnings announcement day in DM and QVS . As a robustness check, I define an alternative measure of the pre-earnings drift (DM_{it}^n) which includes up to n days after t in the return attributed to the announcement itself:

$$DM_{it}^n = 100 \times \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t+n)}} & \text{if } r_{(t,t+n)} > 0 \\ \frac{1+r_{(t-22,t+n)}}{1+r_{(t-22,t-1)}} & \text{if } r_{(t,t+n)} < 0 \end{cases} \quad (15)$$

Figure 7 shows that the time-series trends in DM are similar for choices of n up to 5.

In a similar vein, I define an alternative version of QVS ($QVS_{i,t}^n$) which includes up to n days after t in the volatility attributed to the announcement itself:

$$100 \times \sum_{\tau=-22}^{-1} r_{i,t+\tau}^2 / \sum_{\tau=-22}^n r_{i,t+\tau}^2 \quad (16)$$

Figure 8 shows that the time-series trends in QVS are similar for choices of n up to 5.

These same concerns could also apply to the baseline OLS estimates. Table 14 shows that including up to 5 days after the earnings announcement does not qualitatively change my baseline results for QVS or DM .

C.6 Implied volatility difference

To map the methodology in Kelly et al. (2016) to my setting, I start by identifying all of the regular monthly option expiration dates, which typically occur on the 3rd Friday of each month. Letting τ denote an earnings announcement date, the goal is to identify expiration dates a , b , and c , such that $a < \tau < b < c$. To avoid issues inherent in the calculating implied volatility for short-maturity options (Beber and Brandt, 2006), b is selected so that it is at

Panel A: QVS						
	Include t+1		Include up to t+3		Include up to t+5	
	(1)	(2)	(3)	(4)	(5)	(6)
Passive Ownership	-39.05*** (3.017)	-25.12*** (8.309)	-36.49*** (2.905)	-20.03** (8.073)	-33.81*** (2.750)	-15.91* (8.116)
Observations	430,401	430,401	430,401	430,401	430,401	430,401
R-Squared	0.222	0.233	0.196	0.218	0.174	0.203
Firm + Year/Quarter FE	✓	✓	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value	Equal	Value
Panel B: DM						
	Include t+1		Include up to t+3		Include up to t+5	
	(1)	(2)	(3)	(4)	(5)	(6)
Passive Ownership	-4.073*** (0.680)	-4.133*** (1.212)	-3.394*** (0.720)	-5.177*** (1.131)	-2.946*** (0.813)	-4.682*** (1.429)
Observations	430,401	430,401	430,401	430,401	430,401	430,401
R-Squared	0.217	0.261	0.22	0.25	0.223	0.248
Firm + Year/Quarter FE	✓	✓	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓	✓	✓
Weight	Equal	Value	Equal	Value	Equal	Value

Table 14 Sensitivity of QVS and DM results to including a n -day post-earnings-announcement window. Estimates of β from:

$$Outcome_{i,t}^n = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$$

Where $Outcome^n$ is either $QVS_{i,t}^n$, a version of QVS that includes n days after the earnings announcement in the volatility attributed to the earnings announcement itself or $DM_{i,t}^n$, a version of DM that includes n days after the earnings announcement in the return attributed to the earnings day itself. Controls in $X_{i,t}$ include 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.

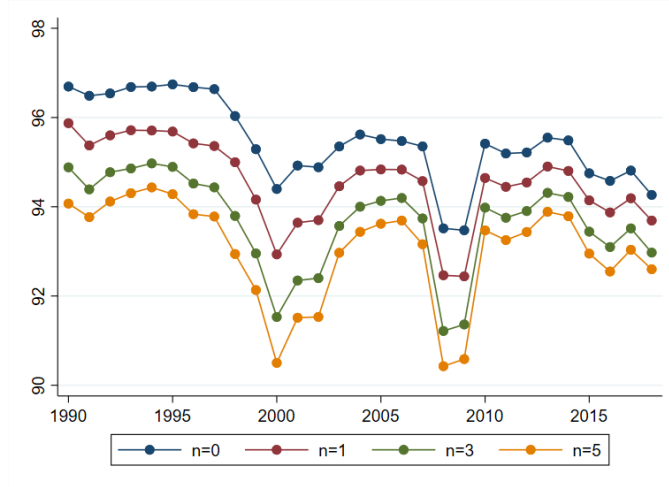


Figure 7. Time series trends in DM^n . This figure plots coefficients from a regression of DM^n on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient.

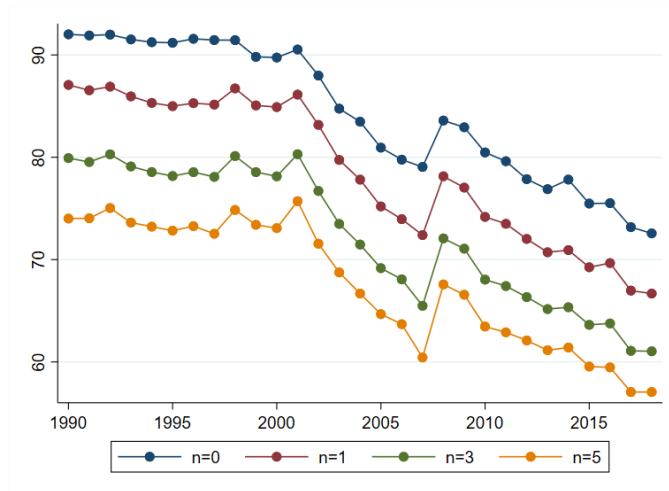


Figure 8. Time series trends in QVS^n . This figure plots coefficients from a regression of QVS^n on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient.

least 5 days after τ .³³

³³This means that if the first regular expiration after the earnings announcement has at least 6 days to maturity at τ , that expiration will be b , and a will be one month before b . If the first regular expiration after the earnings announcement has fewer than 5 days to expiration at τ , b will be the next regular expiration date, and a will be two months before b . c is always chosen to be one month after b .

Having identified a , b , and c , the next step is to compute the average implied volatility associated with each of these expiration dates. For each firm i , on each trading day t , I compute $IV_{i,t,e}$, defined as the equal-weighted average implied volatility across all at-the-money options expiring on date e . Then, I take an equal-weighted average of $IV_{i,t,b}$ over the 20-day window before τ :

$$\overline{IV}_{i,b} = \text{Mean} [IV_{i,(b-s),b} : b-s \in [\tau-20, \tau-1]] \quad (17)$$

$\overline{IV}_{i,a}$ and $\overline{IV}_{i,c}$ are defined analogously, as averages of $IV_{i,t,e}$ over the 20-day windows that end $b-\tau+1$ days before a and c .

The final variable of interest, the implied volatility difference, is defined as:

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2} (\overline{IV}_{i,a} + \overline{IV}_{i,c}) \quad (18)$$

higher values of $IVD_{i,\tau}$ imply that options which span earnings announcements are relatively more expensive.³⁴ The units of IVD are percentage points of implied volatility.

Implied volatility is computed by OptionMetrics and runs from 1996 until the end of my sample. I use the WRDS linking suite to match the OptionMetrics data with CRSP. Following Kelly et al. (2016), I keep all options with positive open interest, and define at-the-money options as those with absolute values of delta between 0.4 to 0.5. For a firm/earnings-announcement pair to be included, it must be that a and b are no more than two months apart, and c is no more than one month after b .³⁵

Figure 9 plots the cross-sectional average of IVD by quarter. Numbers greater than zero are evidence that options which span earnings announcements are more expensive than those with surrounding maturities. Consistent with the increase in earnings-day volatility (i.e., the decline in QVS), on both an equal-weighted and value-weighted basis, IVD has increased by about 5 over the past 25 years.

³⁴One 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 $\tilde{IV}D_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2} (\overline{IV}_{i,a} + \overline{IV}_{i,c})$.

³⁵Suppose firm i has an earnings announcement on 1/5/2021. Then a should be 12/18/2020, b should be 1/15/2021 and c should be 2/19/2021. Suppose, however, that between 1/21/2021 and 2/10/2021 there are no options expiring on 2/19/2021 with positive open interest and absolute values of delta between 0.4 and 0.5. This last filter prevents e.g., the use of options expiring 3/19/2021 in place of options expiring 2/19/2021 to compute $\overline{IV}_{i,c}$.



Figure 9. Time-series trends in IVD . Equal-weighted and value-weighted averages of IVD by quarter. Red dots represent cross-sectional averages and blue lines represent LOWESS filters with bandwidths equal to 20% of quarters in the dataset.

C.7 Alternative explanations for the decline of price informativeness

In this subsection, I discuss three threats to identification in my baseline regressions (1) Regulation Fair Disclosure (2) the rise of algorithmic trading and (3) the relationship between passive ownership and corporate governance.

C.7.1 Regulation Fair Disclosure (Reg FD)

Before Reg FD was passed in August, 2000, firms would disclose earnings information to selected analysts before it became public. This information likely made its way into prices before it was formally announced, increasing pre-earnings announcement price informativeness. After Reg FD passed, firms were no longer allowed to selectively disclose material information, and instead must release it to all investors at the same time.

Reg FD could be driving the trends in QVS and DM , as there was a large negative shock to the amount of information firms released before earnings announcements after it was passed. QVS and DM , however, continue to trend in the same direction after Reg

FD was implemented. Reg FD could still explain these results if the value of information received by analysts before Reg FD decayed slowly. While this is possible, my prior is that information obtained in 2000 would not be relevant for more than a few years.

For Reg FD to be driving the cross-sectional relationship between passive ownership and pre-earnings price informativeness, it would have to disproportionately affect firms with high passive ownership. This is because all the regressions have year-quarter fixed effects, which should account for any level shifts in price informativeness after Reg FD was passed. To further rule out this channel, in the main body of the paper, I re-run the cross-sectional regressions using only post-2000 data. The point estimates are quantitatively similar, which alleviates concerns that my results being driven by Reg FD.

Another possibility is that Reg FD changed the way insiders (e.g., directors or senior officers) behaved, or led to changes in the enforcement of insider trading laws (Coffee, 2007). If this were purely a time-series effect, however, it cannot be driving the OLS regressions which have time fixed-effects. To further rule out the insider behavior channel, I used the Thompson Insiders data to compute insider buys/sells as a percent of total shares outstanding for each firm in my dataset.

In terms of basic properties, insider buys and sells have been decreasing since the mid-1990's. Both average annual buys and sells went down slightly more for stocks with more passive ownership, but this effect is only weakly statistically significant. I then examined insider buys/sells in 22-day windows before/after earnings announcements. Both buys and sells have decreased before and after earnings announcements, broadly following the trend toward decreased insider activity. There is no statistically significant relationship, however, between passive ownership and insider buys/sells before or after earnings announcements. This is at least suggestive evidence that changes in insider behaviour is not driving my OLS estimates.

C.7.2 The Rise of algorithmic trading (AT) activity

Weller (2018) shows that Algorithmic Trading (AT) activity is negatively correlated with pre-earnings price informativeness. The proposed mechanism is algorithmic traders back-run informed traders, reducing the returns to gathering firm-specific fundamental information. AT activity increased significantly over my sample period, and could be responsible for some of the trend toward decreased average pre-earnings price informativeness.

It is difficult to measure the role of algorithmic traders in the trends toward decreased pre-earnings price informativeness, as I only have AT activity proxies between 2012-2018. I can, however, measure the effect of AT activity on the cross-sectional regression results. For AT activity to influence the regression estimates, it would have to be correlated with passive ownership, which I find plausible because: (1) Passive ownership is higher in large, liquid stocks, where most AT activity occurs. This, however, should not affect my results, as I condition on firm size in all the cross-sectional regressions and (2) High ETF ownership will attract algorithmic traders implementing ETF arbitrage. The effect of time trends in AT activity should be absorbed by the time fixed effects.

To rule out this channel, I construct the 4 measures of AT activity used in Weller (2018) from the SEC MIDAS data. MIDAS has daily data for all stocks traded on 13 national exchanges from 2012 to present. The AT measures are (1) odd lot ratio, (2) trade-to-order ratio, (3) cancel-to-trade ratio and (4) average trade size. Measures 1 and 3 are positively correlated with AT activity, while the opposite is true for measures 2 and 4. Consistent with Weller (2018), I (1) Truncate each of the AT activity variables at the 1% and 99% level by year to minimize the effect of reporting errors (2) calculate a moving average for each of these measures in the 21 days leading up to each earnings announcement and (3) take logs to reduce heavy right-skewness. Only 1% of MIDAS data cannot be matched to CRSP, so the drop in sample size relative to the baseline OLS regressions is almost entirely the result of restricting to data between 2012 and 2018.

Table 3 in the main body of the paper adds the 4 AT activity measures to the right-hand side of the baseline OLS regressions. The point estimates are not significantly changed by including these controls, suggesting that the correlation between passive ownership and AT activity is not driving my results.

C.7.3 Effect of passive ownership on corporate governance

Given the literature on the effects of passive owners on corporate governance (Appel et al., 2016), one could worry that passive ownership's primary effect is to change governance, and then governance changes price informativeness. One mechanism would be that better governance leads to fewer information leaks, which in turn makes prices less informative

before earnings announcements.³⁶

To test this, I quantify corporate governance using the entrenchment index (E index) of Bebchuk et al. (2009). Using data from ISS between 1990 and 2018, I calculate this as the sum of indicator variables for the presence of: (1) a staggered (classified) board (2) a limitation on amending bylaws (3) a limitation on amending the corporate charter (4) a requirement of a supermajority to approve a merger (5) golden parachutes for management/board members and (6) a poison pill.³⁷ I then run a regression of the E index on passive ownership. Given that the E index is only defined annually, I use end of year data for passive ownership as well as all the control variables. Also, given that ISS coverage is not equally spread across the firm size distribution, I do not report the value-weighted regression results.

Consistent with Gloßner (2018), Table 15 shows there isn't a statistically significant relationship between governance and passive ownership. The effect of a 15% increase in passive ownership on the E index is less than 0.05, so the effect of passive ownership on governance is also economically small relative both to the mean (≈ 3) and the standard deviation (≈ 1.5). I also find that my baseline regressions are unchanged by explicitly controlling for the E index. Jointly, this evidence suggests that the relationship between passive ownership and corporate governance is not driving my results.

D Causal analysis

D.1 Firm size and passive ownership

As discussed in the introduction, a possible threat to identification is the relationship between passive ownership and firm size. Figure 10 plots the relationship between the passive ownership share and the percentile of market capitalization for observations in December 2018. The relationship is positive with a univariate R-squared of 25%. For very large stocks (those in the top 20% of market capitalization) the relationship starts to break down and invert. One explanation for this is that mid-cap indices (e.g., the Russell 2000) have

³⁶There is, however, mixed evidence on the relationship between passive ownership and corporate governance. For example, quoting Gloßner (2018), “I also find that passive investors have no significant effect on corporate social responsibility (CSR) ...”, and the measure of CSR he uses includes corporate governance.

³⁷Data from 1990-2006 is in a separate database – “ISS – Governance Legacy” – than the data from 2007 onward.

	(1)	(2)	(3)
Passive Ownership	0.402 (0.380)	0.581 (0.427)	0.212 (0.444)
Observations	43,221	39,937	39,937
R-Squared	0.838	0.839	0.839
Firm + Year/Quarter FE	✓	✓	✓
Matched to Controls		✓	✓
Firm-Level Controls			✓
Weight	Equal	Equal	Equal

Table 15 Passive ownership and entrenchment. Table with estimates of β from:

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

where $EIndex$ is the entrenchment index of Bebchuk et al. (2009). Controls in $X_{i,t}$ include 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.

a relatively larger passive ownership share than large-cap indices (e.g. the Russell 1000) (Pavlova and Sikorskaya, 2022).

D.2 Russell Details

I use the following procedure, based on Chang et al. (2015) and Coles et al. (2022), to compute the proxy for Russell's May market capitalization ranks. I also incorporate the improvement from Ben-David et al. (2019), which accounts for the exact day Russell rebalances the indices:

- Compute the number of shares outstanding/market capitalization on the index rebalancing date according to CRSP. To do this, start with the CRSP daily security file. Merge this with the list of dates from Ben-David et al. (2019) to identify the trading date closest to the Russell index rebalancing date.
 - An adjustment has to be made if a PERMCO (permanent company identifier in CRSP) has multiple associated PERMNOs (permanent security identifier in

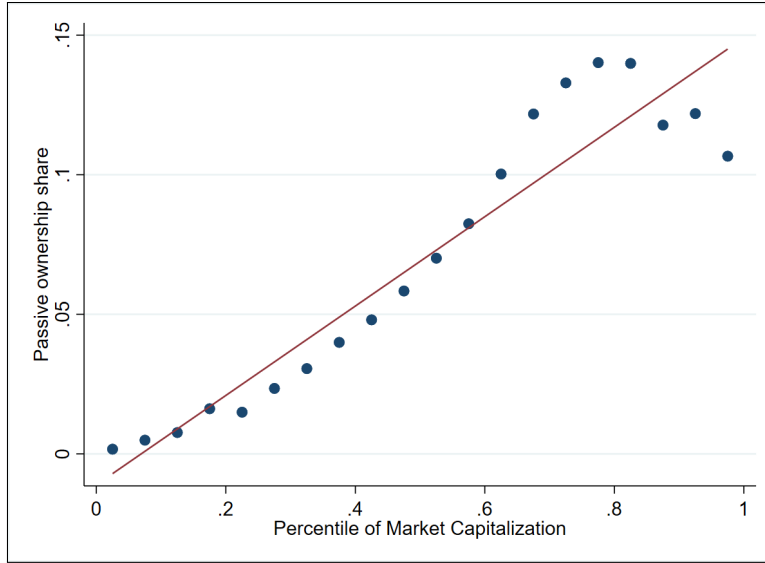


Figure 10. Passive ownership and percentile of market capitalization. Data from 12/2018. Includes all firms with both non-missing passive ownership and non-missing market capitalization.

CRSP). There are two broad cases to consider: (1) If only one of the PERMNOs is in the Russell 3000 universe, for each PERMNO, compute total market capitalization at the PERMCO level (2) If more than one of the PERMNOs is in the Russell 3000 universe, compute the market capitalization for each PERMNO individually.³⁸

- Use the raw Compustat data to identify the release date of quarterly earnings (RDQ). If this is missing, follow the procedure in Chang et al. (2015). Specifically, if the missing RDQ is associated with a fiscal year end (10K):
 - If the fiscal year end is before 2003, set RDQ to 90 days after the period end date.
 - If the fiscal year end is between 2003 and 2006, and the firm has a market capitalization greater than 75 million, set RDQ to 75 days after the period end date. If the firm has a market cap less than 75 million, set RDQ to 90 days after the period end date.
 - If the fiscal year end is 2007 or later, and the firm has a market capitalization great than 700 million, set RDQ to 60 days after the period end date. If the firm

³⁸I would like to thank Simon Gloßner for bringing this to my attention (Gloßner, 2018).

has a market capitalization between 75 and 700 million set RDQ to 75 days after the period end date. Finally, if the firm has a market capitalization less than 75 million, set RDQ to 90 days after the period end date.

If the missing RDQ is associated with a fiscal quarter end (10Q):

- If the fiscal year-quarter is before 2003, set RDQ to 40 days after the end of the fiscal period.
 - If the fiscal year-quarter is in or after 2003, and the firm has a market capitalization of more than 75 million, set RDQ to 40 days after the fiscal quarter end. If the firm has a market capitalization smaller than 75 million, set RDQ to 45 days after the fiscal quarter end.
- Compute the number of shares outstanding on the index rebalancing date according to the Compustat data. Start with the number of shares outstanding in Compustat (CSHOQ). Then, adjust for changes in the number of shares outstanding between the release date of earnings information (RDQ), and the Russell index rebalancing date. To do this, start at RDQ, and apply all of the CRSP factor to adjust shares between RDQ and the rebalancing date.
 - Map the Russell index member data to CRSP using the following procedure:
 - First, create a new CUSIP variable that is equal to historical CUSIP if that is not missing, and is equal to current CUSIP otherwise. Merge on this new CUSIP variable and date.
 - For the remaining unmatched firms, merge on ticker, exchange and date.
 - For the remaining unmatched firms that had non-missing historical CUSIP, but weren't matched on historical CUSIP to the Russell data, merge on current CUSIP and date.
 - For the remaining unmatched firms, merge on ticker and date. Note that in some of these observations, the wrong field is populated (e.g., the actual ticker was put into the CUSIP field in the Russell data), so that needs to be fixed before doing this last merge.
 - Merge CRSP and Compustat using the CRSP/Compustat merged data.
 - Use the following procedure to compute May market capitalization: If the shares outstanding from the Compustat data is larger than the shares outstanding from CRSP,

use that number of shares outstanding to compute market capitalization. Otherwise, use the shares outstanding in the CRSP data to compute market capitalization. In either case, compute market capitalization using the closing price on the day closest to the index rebalancing date.

With this May market capitalization proxy, I use the following procedure, also based on Coles et al. (2022) to predict index membership and identify the cohorts of treated/control firms:

- Each May, rank stocks by market capitalization.
- Identify the 1000th ranked stock, and compute the bands as $\pm 2.5\%$ of the total market capitalization of the Russell 3000.³⁹
- Identify the cutoff stocks at the top and bottom bands. For stocks switching to the 2000, this will be the first stock that is ranked below the lower band. For stocks switching to the 1000, this will be the first stock that is ranked above the upper band.
- The cohorts of treated/control firms are those within ± 100 ranks around these cutoff stocks. For the possible switchers to the 2000, they must have been in the 1000 the previous year, while for possible switchers to the 1000, they must have been in the 2000 the previous year.
- If a firm was in the 1000 last year, as long it has a rank higher than the cutoff, it will stay in the 1000. If a firm was in the 2000 last year, as long as it has a rank lower than the cutoff, it will stay in the 2000. Otherwise, the firm switches.
 - When using this data, to identify actual switchers, it is easy to miss that in 2013, Russell records the rebalancing in July, rather than June

D.3 Alternative instruments

D.3.1 Moving from the Russell 2000 to the Russell 1000

As discussed in the main body of the text, firms experience a mechanical decrease in passive ownership after they are moved from the Russell 2000 to the Russell 1000. This is

³⁹In reality, the bands are $\pm 2.5\%$ of the Russell 3000E, not the Russell 3000. The data I have from FTSE Russell only has Russell 3000 firms, which is why I use that instead. I discussed this with the authors of Coles et al. (2022) and they find using the total market capitalization of the 3000 vs. 3000E makes almost no difference to the accuracy of predicted index membership.

because (1) they go from being the largest firm in a value-weighted index of small firms, to the smallest firm in a value-weighted index of large firms and (2) the passive ownership share is higher for the Russell 2000 than the Russell 1000 (Pavlova and Sikorskaya, 2022). This index change, therefore, seems like a natural instrument for passive ownership.

Again, following Coles et al. (2022), I choose the control firms to be those within ± 100 ranks of the upper band that were in the Russell 2000 the previous year. Figure 11 shows the problem with this IV: the change in passive ownership associated with switching from the 2000 to the 1000 is small and temporary. Within 12 months of switching, passive ownership is almost back at the pre index-rebalancing level.

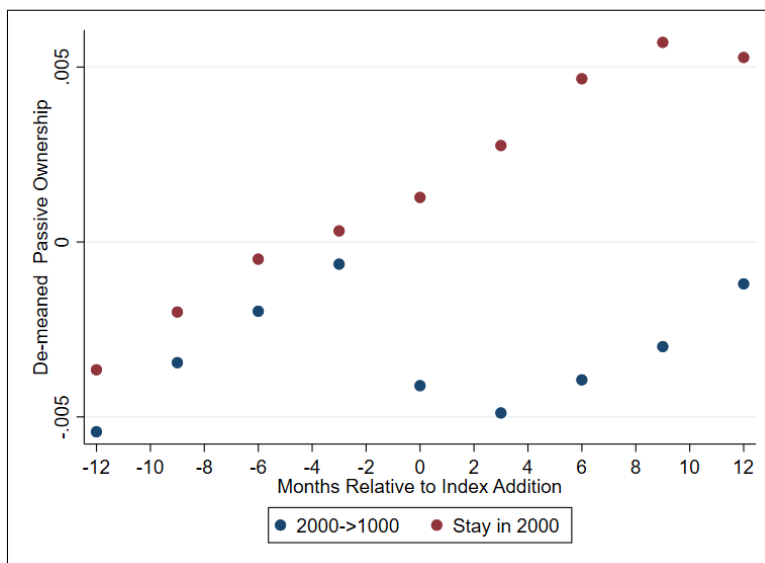


Figure 11. Russell 1000/2000 Reconstitution and Changes in Passive Ownership. Average level of passive ownership for firms that stay in the Russell 2000 (control firms) and firms that moved from the Russell 2000 to the Russell 1000 (treated firms). Passive ownership is demeaned within each cohort.

D.3.2 Blackrock's acquisition of Barclays Global Investors

Another possible instrument for passive ownership can be constructed around Blackrock's acquisition of Barclays' iShares ETF business in December 2009. This is not an ideal setting for testing my hypothesis because: (1) My proposed mechanism has no predictions for the effects of increased concentration of ownership among passive investors (Azar et al. (2018),

Massa et al. (2021)) and (2) While there may have been a *relative* increases in flows to iShares ETFs, compared to all other ETFs (Zou, 2018), I do not find a significant increase in overall ETF ownership for the stocks owned by iShares funds. Given that my right-hand side variable of interest is the percent of shares owned by passive funds, my proposed mechanism has no predictions for the effect of moving dollars from iShares ETFs to non-iShares ETFs.

D.4 Statistical significance of instrumental variables vs. reduced form

In Table 4, the IV regressions are highly significant, while the reduced form regressions are insignificant. The concern is that, as discussed in Chernozhukov and Hansen (2008b), a significant IV with an insignificant reduced form potentially indicates a weak instruments problem. In their notation:

$$\begin{aligned}\text{Structural} & : y = X\beta + \varepsilon \\ \text{First stage} & : X = Z\Pi + V \\ \text{Reduced Form} & : Y = Z\gamma + U\end{aligned}$$

Specifically, suppose the instruments are weak so $cov(Z, X)$ is close to zero. Then $(Z'Y) / (Z'X)$ i.e., the IV estimate of β might be large, but not because the true β is large. They argue that another way to test whether the true $\beta = 0$ is to check if $\gamma = 0$ i.e., test whether the reduced form is insignificant.⁴⁰ At a high level, this is likely not a problem in my setting, as the first stage is very strong ($F > 300$ for the S&P experiment and $F > 200$ for the Russell experiment).

Further, as pointed out by Lochner and Moretti (2004), for a given IV standard error, the reduced form standard errors can be arbitrarily large or small. To formalize the argument,

⁴⁰This may be an indication that $\beta = 0$ because $\gamma = \beta \cdot \Pi$ i.e., if $\beta = 0$ then γ will be zero.

consider the case of a univariate structural regression and a single instrument. The model is

$$\text{Structural} : y_i = \beta x_i + \varepsilon_i \quad (19)$$

$$\text{First stage} : x_i = \gamma z_i + u_i \quad (20)$$

$$\text{Reduced Form} : y_i = \overbrace{\beta\gamma}^{\alpha} z_i + \overbrace{\beta u_i + \varepsilon_i}^{v_i} \quad (21)$$

For simplicity, assume all variables are mean zero and have *iid* sampling so a standard law of large numbers and central limit theorem hold. Further, assume that $E[z_i \varepsilon_i] = 0$, but $E[x_i \varepsilon_i] = E[u_i \varepsilon_i] \neq 0$. This is the exclusion restriction i.e., the assumption that the instrument z_i cannot be correlated with ε_i , which is why $E[z_i \varepsilon_i] = 0$. The exclusion restriction also implies $E[x_i \varepsilon_i] = E[(\gamma z_i + u_i) \varepsilon_i] = E[u_i \varepsilon_i]$

Under these assumptions, the usual IV results still hold, namely that the OLS is inconsistent and the IV and reduced form are consistent. Writing out the definition of the OLS estimator:

$$\hat{\beta}_{OLS} = \frac{N^{-1} \sum_i y_i x_i}{N^{-1} \sum_i x_i^2} = \beta + \frac{N^{-1} \sum_i (\gamma z_i + u_i) \varepsilon_i}{N^{-1} \sum_i x_i^2}$$

$\hat{\beta}_{OLS}$ does not converge in probability to β (i.e., the true beta) because of the correlation between x_i and ε_i :

$$\hat{\beta}_{OLS} - \beta \xrightarrow{p} \frac{E[u_i \varepsilon_i]}{E[x_i^2]} \neq 0.$$

Writing out the definition of the IV estimator:

$$\hat{\beta}_{IV} = \frac{N^{-1} \sum_i y_i z_i}{N^{-1} \sum_i x_i z_i} = \beta + \frac{N^{-1} \sum_i \varepsilon_i z_i}{N^{-1} \sum_i x_i z_i}$$

Unlike the OLS estimator, $\hat{\beta}_{IV}$ will converge in probability to the true β because the exclusion restriction implies $E[\varepsilon_i z_i] = 0$. The distribution of the IV estimator is:

$$\sqrt{N}(\hat{\beta}_{IV} - \beta) = \frac{\frac{1}{\sqrt{N}} \sum_i \varepsilon_i z_i}{N^{-1} \sum_i x_i z_i} \xrightarrow{d} N\left(0, \frac{E[\varepsilon_i^2 z_i^2]}{(E[x_i z_i])^2}\right).$$

Finally, writing out the definition of the reduced form estimator:

$$\hat{\alpha}_{RF} = \frac{N^{-1} \sum_i y_i z_i}{N^{-1} \sum_i z_i^2}, = \alpha + \frac{N^{-1} \sum_i v_i z_i}{N^{-1} \sum_i z_i^2}$$

Like the IV estimator, $\hat{\alpha}_{RF}$ will converge in probability to the true α because, by construction, $E[v_i z_i] = 0$. The distribution of the reduced form estimator is:

$$\sqrt{N}(\hat{\alpha}_{RF} - \alpha) = \frac{\frac{1}{\sqrt{N}} \sum_i v_i z_i}{N^{-1} \sum_i z_i^2} \xrightarrow{d} N \left(0, \frac{E[v_i^2 z_i^2]}{(E[z_i^2])^2} \right).$$

Assuming homoskedasticity, the distribution of the centered t-statistics for the IV and reduced form estimators are:

$$t_{\hat{\beta}_{IV}} = \frac{\left(\frac{\sum_i y_i z_i}{\sum_i x_i z_i} - \beta \right)}{\sqrt{\frac{(N^{-1} \sum_i \hat{\varepsilon}_i^2)(\sum_i z_i^2)}{(\sum_i x_i z_i)^2}}} = \frac{\frac{1}{\sqrt{N}} \sum_i z_i \varepsilon_i}{\sqrt{(N^{-1} \sum_i \hat{\varepsilon}_i^2)(N^{-1} \sum_i z_i^2)}} \xrightarrow{d} N(0, 1)$$

And

$$t_{\hat{\alpha}_{RF}} = \frac{\left(\frac{\sum_i y_i z_i}{\sum_i z_i^2} - \alpha \right)}{\sqrt{\frac{N^{-1} \sum_i \hat{v}_i^2}{\sum_i z_i^2}}} = \frac{\frac{1}{\sqrt{N}} \sum_i v_i z_i}{\sqrt{(N^{-1} \sum_i \hat{v}_i^2)(N^{-1} \sum_i z_i^2)}} \xrightarrow{d} N(0, 1).$$

Thus, their joint distribution is:

$$\begin{bmatrix} t_{\hat{\alpha}_{RF}} \\ t_{\hat{\beta}_{IV}} \end{bmatrix} \rightarrow N \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & \rho_{\varepsilon, v} \\ \rho_{\varepsilon, v} & 1 \end{bmatrix} \right)$$

where

$$\rho_{\varepsilon, v} = \frac{E[v_i \varepsilon_i]}{\sqrt{(E[v_i^2])(E[\varepsilon_i^2])}} = \frac{\beta \frac{E[u_i \varepsilon_i]}{E[\varepsilon_i^2]} + 1}{\sqrt{\beta^2 \frac{E[u_i^2]}{E[\varepsilon_i^2]} + 2\beta \frac{E[u_i \varepsilon_i]}{E[\varepsilon_i^2]} + 1}} \quad (22)$$

Equation 22 implies that if the true $\beta = 0$, $\rho_{\varepsilon, v}$ will be equal to 1 and the t-statistics will be perfectly correlated asymptotically. Alternatively, if $\beta \neq 0$, then $\rho_{\varepsilon, v}$ will be less than 1 and these two t-statistics will not be perfectly correlated, even asymptotically. Thus, it is possible to have a significant IV estimate and insignificant reduced form estimate and this

becomes more likely as $\rho_{\varepsilon, v}$ decreases.

Empirically, the econometrician does not know α and β , so one cannot compute the centered t -statistics. Instead, following Lochner and Moretti (2004) and computing these t -statistics under the $\alpha = \beta = 0$ null yields:

$$\begin{aligned} t_{\hat{\beta}_{IV}} &= \frac{\sum_i y_i z_i}{\sqrt{(N^{-1} \sum_i \hat{\varepsilon}_i^2) (\sum_i z_i^2)}} \\ t_{\hat{\alpha}_{RF}} &= \frac{\sum_i y_i z_i}{\sqrt{(N^{-1} \sum_i \hat{v}_i^2) (\sum_i z_i^2)}} \end{aligned}$$

and taking their ratio yields:

$$\frac{t_{\hat{\beta}_{IV}}}{t_{\hat{\alpha}_{RF}}} = \frac{\sqrt{(N^{-1} \sum_i \hat{v}_i^2)}}{\sqrt{(N^{-1} \sum_i \hat{\varepsilon}_i^2)}} \xrightarrow{p} \sqrt{\frac{E[v_i^2]}{E[\varepsilon_i^2]}} = \sqrt{\frac{\beta^2 E[u_i^2] + 2\beta E[u_i \varepsilon_i] + E[\varepsilon_i^2]}{E[\varepsilon_i^2]}} \quad (23)$$

Thus, under the $\beta = \alpha = 0$ null, these t -statistics will be perfectly correlated asymptotically.⁴¹

Empirically, in Table 4, I am testing whether $\hat{\beta}_{IV} = 0$ and $\hat{\alpha}_{RF} = 0$. With this in mind, assuming the true β and α are not zero, there are three things to consider:

1. Equation 23 shows that even fixing the IV t -statistic, the standard errors in the reduced form can be arbitrarily large or small depending on the correlation between the residuals in the structural and first stage regressions. Suppose, for example, $t_{\hat{\beta}_{IV}} = -2.5$ so the IV is statistically significant, the true β is -0.5, and $E[u_i^2] = E[\varepsilon_i^2] = 1$. Then, if the covariance between u_i and ε_i is lower than -0.4, the reduced form coefficient will not be both negative and significant at the 5% level. This is because with these parameters $\sqrt{\frac{\beta^2 E[u_i^2] + 2\beta E[u_i \varepsilon_i] + E[\varepsilon_i^2]}{E[\varepsilon_i^2]}} = 1.28$ and $-2.5/1.28 > -1.96$.
2. More generally, in my setting I expect $\beta < 0$ i.e., passive ownership decreases price informativeness. If this is the case, as the covariance between u_i and ε_i becomes more negative, we expect the uncentered t -statistic for the IV to be relatively larger than the uncentered t -statistic for the reduced form. This is because this increasing negative

⁴¹If $\beta = \alpha = 0$ are not the true parameters, then the distribution of these t -statistics will not be asymptotically normal. In fact, they will not have a limiting distribution and will tend to diverge as N grows (i.e., the mean of the distribution will become infinitely large in absolute value). These two t -statistics, however, will still be perfectly correlated in large samples.

covariance will tend to increase $\beta E[u_i \varepsilon_i]$ in the numerator of Equation 23, increasing the ratio of the IV t-statistic to the reduced form t-statistic.

3. As the number of observation in my sample grows, we expect both the IV and reduced form t-statistics to increase because this will decrease:

$$\hat{\sigma}_\varepsilon^2 = (y_i - \hat{\beta}_{IV} x_i)'(y_i - \hat{\beta}_{IV} x_i)/N$$

and

$$\hat{\sigma}_v^2 = (y_i - \hat{\alpha}_{RF} z_i)'(y_i - \hat{\alpha}_{RF} z_i)/N$$

Economic mechanism for correlation in error terms: Shadow indexing

The analysis above shows that in my setting the reduced form is more likely to be insignificant if $\beta < 0$ and $Cov(u_i, \varepsilon_i) < 0$. In this section, I argue this is likely to be the case because of shadow indexing, defined as funds or investors which are passive, but don't explicitly say so (e.g., an institutional investor who is internally replicating the S&P 500 index). The logic is that when a firm gets a bigger than expected increase in passive ownership from changing indices (i.e., the first stage residual u_i is positive), the true change in passive ownership is even larger. And, because $\beta < 0$, the structural regression will undershoot the true change in price informativeness (i.e., the structural equation residual ε_i will be negative), leading to a negative correlation between u_i and ε_i .

More formally, suppose true passive ownership, $passive_{i,t}^*$, is equal to ownership by explicitly passive funds, $passive_{i,t}$ (i.e., the measure of passive ownership in the paper), plus ownership by shadow indexers, $shadow_{i,t}$. Suppose further that the data generating process for price informativeness is:

$$informativeness_{i,t} = a_i + \beta passive_{i,t}^* + \varepsilon_{i,t}$$

which implies that true passive ownership is what matters for price informativeness. Now, suppose that when a firm is added to a major index, it may also be added to several sub-indices. For example, when a firm moves from the Russell 1000 to the 2000, it may also be added to the Russell 2000 growth. Finally, suppose that shadow indexing is proportional to observed indexing i.e., $shadow_{i,t} = \psi \cdot passive_{i,t}$ where $\psi > 0$. This might be the case if e.g., there are shadow indexers who also track the sub-indices.

Now, in my IV, I measure the average difference in passive ownership for firms around the cutoff before index rebalancing to estimate the change in passive ownership a firm will

receive from being added to the index, which I call $PassiveGap_{i,t}$. But suppose that firm i also gets added to several sub-indices, so the true increase in passive ownership is larger than $PassiveGap_{i,t}$. Recalling the first stage regression:

$$passive_{i,t} = b \cdot added_{i,t} + c \cdot post_{i,t} + d \cdot (added_{i,t} \times post_{i,t} \times PassiveGap_{i,t}) + u_{i,t}$$

In this case, $u_{i,t}$ would be positive, because firm i received a larger than expected increase in passive ownership because it was also added to the sub-indices.

Further, the true level of price informativeness for this firm would be

$$informativeness_{i,t} = \beta \cdot passive_{i,t}^* + \varepsilon_{i,t}$$

but because I only observe $passive_{i,t}$ this becomes

$$\begin{aligned} informativeness_{i,t} &= \beta \cdot passive_{i,t} + (\varepsilon_{i,t} + \beta \cdot shadow_{i,t}) \\ \Leftrightarrow informativeness_{i,t} &= \beta \cdot passive_{i,t} + \tilde{\varepsilon}_{i,t} \end{aligned}$$

where $\tilde{\varepsilon}_{i,t} = \varepsilon_{i,t} + \beta \cdot shadow_{i,t}$. In this setting $u_{i,t}$ and $\tilde{\varepsilon}_{i,t}$ are going to have negative covariance, because $shadow_{i,t}$ is positively related to $passive_{i,t}$. And if $\beta < 0$, then $\beta E[u_i \varepsilon_i] > 0$, which according to Equation 23 would tend to make the reduced form have a smaller t-statistic than the IV.

Simulation evidence

I use simulations to understand just how large the correlation between u_i and ε_i would need to be to generate a scenario where I fail to reject the null via the reduced form but reject the null via the IV. Specifically, I simulate the setup in Equation 19 (the model with a univariate structural regression and a single instrument), varying the sign and the strength of the correlation between u_i and ε_i . Given that the sample size matters (both $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_v^2$ depend on N), I choose $N = 30,000$ to match the number of observations in Panel A of Table 4. I set $\beta = -0.25$, $\gamma = 0.5$, although all results are similar using any $\beta < 0$ and $\gamma \neq 0$. Finally, to ensure that the IV and reduced form estimates are not statistically significant in every simulation, I add additional noise to the system, scaling all ε by 5 and all u by 10.

Figure 12 plots the fraction of simulations where the t-statistic from the IV is less than

-1.96, but the t-statistic from the RF is greater than -1.96. The first dot on the far left of the plot shows that even if u_i and ε_i are uncorrelated, the RF is less likely to be statistically significant than the IV. This is not surprising, as even if $E[u_i \varepsilon_i] = 0$, the ratio in Equation 23 will be bigger than one.

The blue dots show that as the correlation between u_i and ε increases, the RF becomes more statistically significant on average. This is because the numerator in Equation 23 shrinks, as this positive covariance between u_i and ε is being multiplied by β , which is less than 0. Finally, the red dots show that as the correlation between u_i and ε becomes more negative, the RF becomes even less significant on average than the IV. In this case, the negative covariance between u_i and ε is being multiplied by the negative beta, which increases the numerator of Equation 23.

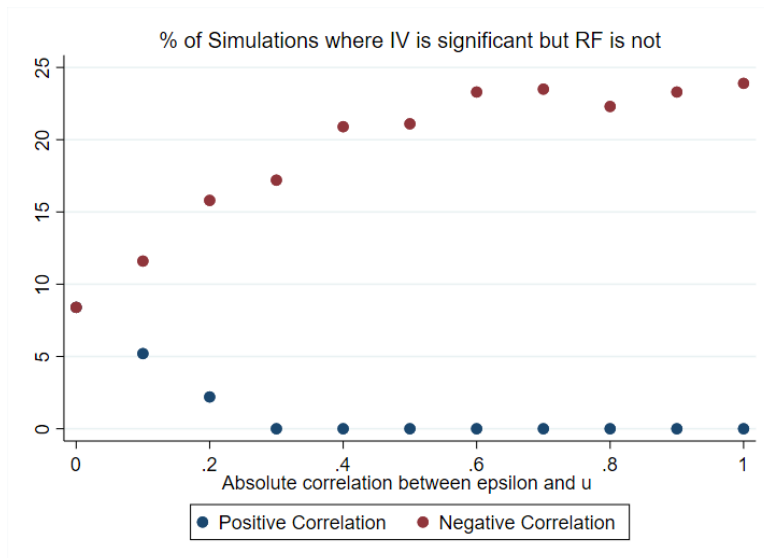


Figure 12. Comparison of statistical significance. Each dot represents the percentage of simulations where the instrumental variables specification is statistically significant, but the reduced form is not. The blue dots are from simulations where ε positively correlated with u , while the red dots are from simulations where ε is negatively correlated with u . Moving from left to right increases the (absolute) correlation between ε and u .

Another explanation is that, as raised in the paper, the reduced form doesn't say anything about the *level* of passive ownership. The reduced form only speaks to changes in passive ownership, but if the level is what truly matters for price informativeness, the reduced form results may be weaker.

D.5 Effect of treatment on total institutional ownership

One concern with the quasi-experimental results is that non-passive institutional ownership may also increase after a firm is added to the S&P 500 or switches from the Russell 1000 to the Russell 2000. This could contaminate my results, as the effects of institutional ownership on a variety of factors that could influence price informativeness are well documented (O’Brien and Bhushan (1990), Asquith et al. (2005), Velury and Jenkins (2006), Chung and Zhang (2011), Aghion et al. (2013)). At a high level, I am not concerned about this for two reasons: (1) Total institutional ownership does not change much around index reconstitution events and (2) All my results survive explicitly accounting for changes in institutional ownership around index reconstitutions.

Previous studies have used the Russell reconstitution as a shock to institutional ownership (Boone and White, 2015). More recent papers, however, have shown that when using the May ranks (which I am doing, following the procedure in Coles et al. (2022)), although there is an increase in passive ownership following Russell index reconstitution events, there little change in overall institutional ownership (Gloßner (2018), Appel et al. (2020))

Gloßner’s results for Russell reconstitutions end in 2006, and I am using reconstitutions from 1990-2018, so to make sure his conclusion also applies in my setting, I expand his results to 2018. To this end, for each cohort, I compute the average level of total institutional ownership for treated and control firms each month relative to the reconstitution. Then, I calculate the total change in institutional ownership between month $t = -6$ and $t = 6$ for the treated firms and subtract the same change for the corresponding control firms (these results, however, are not sensitive to this choice of a ± 6 -month window). I do the same for the S&P 500 index additions, but instead subtract an equal-weighted average of the change in institutional ownership for the two corresponding control groups.

I find that, consistent with Gloßner (2019), for the average firm going from the Russell 1000 to the Russell 2000 over my sample, institutional ownership goes up by 0.36% more for treated firms than control firms, and this difference is not statistically significant. For the S&P 500, the average added firm has an additional 1.78% increase in passive ownership relative to the control firms. This difference is statistically significantly different from zero, even though there is a lot of variation from year to year.

I have two additional pieces of evidence to address the concern that total institutional ownership, rather than passive ownership, is driving my results: In the cross-sectional OLS

regressions, I can and do explicitly control for total institutional ownership. In fact, I find there is significant cross-sectional variation in passive ownership within various levels of institutional ownership. For example, Figure 13 plots passive ownership against institutional ownership in 12/2018. These two quantities are positively correlated, with a univariate R-squared of about 50%. This high correlation, however, is to be expected because passive ownership is included in total institutional ownership.

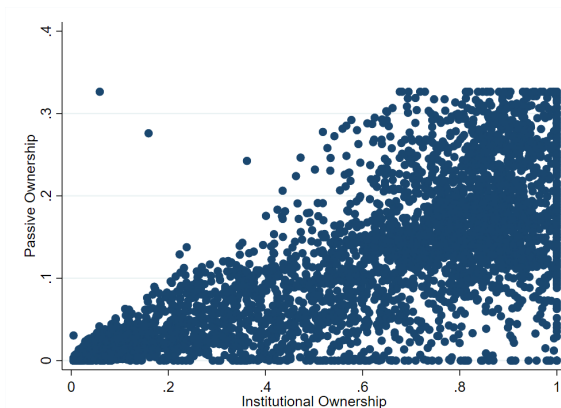


Figure 13. Passive Ownership vs. Institutional Ownership. Plot of passive ownership against total institutional ownership in 12/2018. Both quantities are Winsorized at the 1% and 99% level.

The second piece of evidence is in Table 16, where I replicate all the instrumental variables and reduced form regressions, including total institutional ownership on the right hand side. All the results are quantitatively unchanged from Table 4 in the main body of the paper.

D.6 Alternative instrument: Interaction between CAPM beta and market return

To further allay concerns about my original IV, I have implemented a separate design that uses a different form of identifying variation. My original IV uses time-series variation because it compares differences in informativeness before and after index switching. Alternatively, I leverage cross-sectional variation in index changes that are generated by broad market movements in a small window just before index membership is decided. Specifically, I build on the logic in Bernstein (2015) and use the interaction between a firm's CAPM beta

Panel A: Russell Rebalancing					
	First Stage	QVS IV	RF	DM IV	RF
Post x Treated x Passive Gap Passive Ownership	1.851*** (0.166)	-88.92*** (19.93)	-65.21 (58.91)	-15.06*** (3.61)	-18.61 (11.37)
Observations	30,967	30,967	30,967	30,967	30,967
F-statistic	185				
Panel B: S&P 500 Additions					
	First Stage	QVS IV	RF	DM IV	RF
Post x Treated x Passive Gap Passive Ownership	0.547*** (0.039)	-160.14*** (17.97)	-25.63 (17.86)	-19.83*** (5.47)	-3.77 (4.12)
Observations	185,324	185,324	185,324	185,324	185,324
F-statistic	439				
Cross-sectional regression estimate		-39.48	-39.48	-4.78	-4.78

Table 16 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness (conditioning on total institutional ownership). Estimates from:

$$Passive_{i,t} = \alpha + \beta_1 Post_{i,t} + \beta_2 Passive\ Gap_{i,t} \times Treated_{i,t} \times Post_{i,t} + \gamma Inst. Own_{i,t} + FE + \epsilon_{i,t}$$

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

$$Outcome_{i,t} = \alpha + \beta_4 Post_{i,t} + \beta_5 Passive\ Gap_{i,t} \times Treated_{i,t} \times Post_{i,t} + \gamma Inst. Own_{i,t} + FE + \epsilon_{i,t}$$

where $Outcome_{i,t}$ is QVS or DM and $Post_{i,t}$ is an indicator for observations after the index change. $Passive\ Gap_{i,t}$ is the expected change in passive ownership from being treated. Column 1 in each panel is a first-stage regression. Columns 2 and 4 are instrumental variables regressions. Columns 3 and 5 are reduced-form regressions. Panel A contains observations from Russell rebalancing, while Panel B contains observations from S&P 500 additions. FE are fixed effects for each cohort. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

at the end of March and the cumulative market return from the start of April to the Russell ranking date in May to instrument for passive ownership from July to the following March. For example, in 2010, I use the interaction of CAPM beta and the cumulative market return from April 1st to May 28th to instrument for passive ownership between 7/2010 and 3/2011.

Crucially, the IV regression includes dummy variables for deciles of firm size, formed at the end of March, interacted with year dummies. With these fixed effects, the instrument is leveraging the fact that firms which are similar in size in March, but have differential exposure to market returns from April to late May (based on their CAPM beta) will end up in different indices for index families that rebalance around the end of June (e.g., Russell and S&P). This alternative instrumentation approach is useful because it does not condition on future index membership and because it exploits a different source of variation than the IVs in the main body of the paper (cross-sectional vs. time series).

In this setting, the exclusion restriction is that a firm's CAPM beta times the market return from April to May is exogenous to price informativeness in the year following July. This assumption would be less plausible if stocks with high beta also have high idiosyncratic volatility. To partially address this concern, I explicitly control for idiosyncratic volatility, computed over the same period used to compute CAPM beta.

Column 1 of Table 17 shows the first stage. The relationship between the instrument and passive ownership is positive, consistent with the positive relationship between passive ownership and firm size documented in Figure 10.⁴² Further, the F statistic of 26 suggests the instrument is not weak (Stock and Yogo, 2002). Columns 2 and 4 are the IV regressions, which show point estimates similar in magnitude to the IVs in the main body of the paper. Columns 3 and 5 present the reduced form regressions, which are also negative and statistically significant.

⁴²Figure 10 also shows that this relationship between passive ownership and market capitalization breaks down for very large firms. To this end, in this IV design I exclude firms that were in the S&P 500 at the end of March, but results are quantitatively unchanged by including these observations.

		QVS		DM	
	First Stage	IV	RF	IV	RF
	(1)	(2)	(3)	(4)	(5)
$\beta \times$ market return	0.0365*** (0.007)		-7.070*** (1.679)		-3.812*** (1.044)
Idio. Vol.	-0.220*** (0.017)	-61.60*** (9.786)	-18.90*** (6.037)	-88.67*** (5.655)	-65.65*** (3.490)
Passive Ownership		-193.9*** (43.670)		-104.5*** (26.980)	
Observations	279,827	279,827	279,827	279,827	279,827
F		26.14		26.14	

Table 17 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness (alternative instrument). Estimates from:

$$Passive_{i,t} = \alpha + \beta_1 \beta_{i, March(t)} \times r_{m, April(t) \rightarrow May(t)} + \beta_2 Idio. Vol. + FE + \epsilon_{i,t}$$

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

$$Outcome_{i,t} = \alpha + \beta_4 \beta_{i, March(t)} \times r_{m, April(t) \rightarrow May(t)} + \beta_5 Idio. Vol. + FE + \epsilon_{i,t}$$

where $Outcome_{i,t}$ is *QVS* or *DM*, $\beta_{i, March(t)}$ is firm *i*'s CAPM beta at the end of March and $r_{m, April(t) \rightarrow May(t)}$ is the market return from the start of April to the Russell ranking date in May. Column 1 is a first-stage regression. Columns 2 and 4 are instrumental variables regressions. Columns 3 and 5 are reduced-form regressions. *FE* are fixed effects for the interaction between dummy variables for deciles of market capitalization, formed at the end of March, and dummy variables for each year. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

E Mechanisms Details

E.1 Trends in Earnings Responses

To measure trends in earnings responses, I run the following regression, built on Kothari and Sloan (1992):

$$100 \times r_{i,t} = \alpha + \beta SUE_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \quad (24)$$

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. $r_{i,t}$ is Winsorized at the 1% and 99% level by year. $SUE_{i,t} = \frac{E_{i,t} - E_{i,t-4}}{\sigma_{(t-1, t-8)}(E_{i,t} - E_{i,t-4})}$ where $E_{i,t}$ is earnings-per-share from the IBES unadjusted detail file i.e., “street” earnings, so 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. 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.

Motivated by the asymmetries documented in Figure 6 and Table 13, I also design an earnings-response regression which allows for different reactions to positive and negative surprises:

$$100 \times r_{i,t} = \alpha + \beta_p 1_{SUE_{i,t} \geq 0} \times SUE_{i,t} + \beta_n 1_{SUE_{i,t} < 0} \times |SUE_{i,t}| + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \quad (25)$$

I run regressions 24 and 25 in 5-year rolling windows and plot the β 's in Figure 14. Over the past 30 years, earnings responses have increased by a factor of over $3\times$. Most of this increase was driven by increased responsiveness to SUE s greater than zero. In recent years, however, this trend has reversed, with the response to positive news decreasing and the response to negative news increasing.

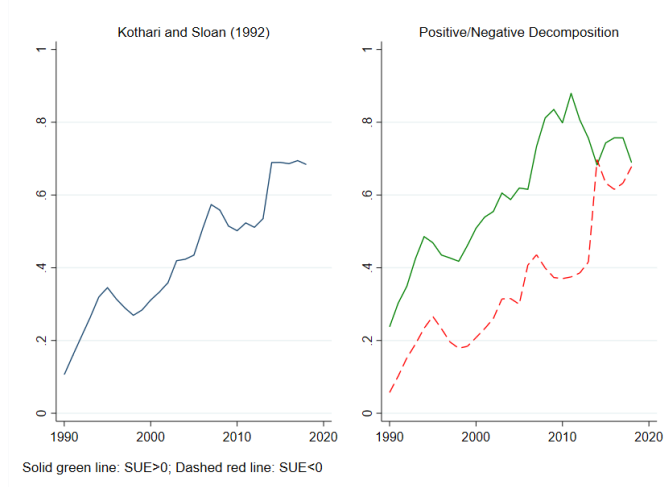


Figure 14. Trends in Earnings Response. Left panel has estimates of β from:

$$100 \times r_{i,t} = \alpha + \beta SUE_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}$$

run in 5-year rolling windows. Right panel has estimates of β_1 and β_2 from Equation 25 i.e., breaking SUE into positive and negative components. Controls in $X_{i,t}$ include 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 regressions contain year-quarter fixed effects, ϕ_t and firm fixed effects ψ_i .

E.2 Robustness of earnings response results

One concern with the earnings-response results is that they are specific to only including the return on the effective earnings announcement date itself on the left-hand side. To alleviate this concern, I run the following regression

$$100 \times r_{i,(t,t+n)} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 (Sys.SUE_{i,t} \times Passive_{i,t}) + \gamma_2 (Idio.SUE_{i,t} \times Passive_{i,t}) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \quad (26)$$

where $r_{i,(t,t+n)}$ is the cumulative log market-adjusted return (in percentage points) from the effective earnings announcement date to $t+n$, Winsorized at the 1% and 99% level by year.

Table 18 shows that even including up to 5 days in $r_{i,(t,t+n)}$ does not change that high passive stocks are especially responsive to idiosyncratic news and that this is robust to using value weights or equal weights.

Panel A: Equal weights					
	n=1	n=2	n=3	n=4	n=5
<i>Sys.SUE</i> × <i>Passive</i>	0.79 (1.152)	1.251 (1.183)	1.496 (1.388)	1.739 (1.560)	1.773 (1.623)
<i>Idio.SUE</i> × <i>Passive</i>	2.729*** (0.239)	2.625*** (0.297)	2.682*** (0.325)	2.744*** (0.345)	2.782*** (0.371)
Observations	333,875	333,875	333,875	333,875	333,875
R-squared	0.063	0.066	0.067	0.066	0.067
Panel B: Value weights					
	n=1	n=2	n=3	n=4	n=5
<i>Sys.SUE</i> × <i>Passive</i>	0.28 (2.214)	-0.397 (2.743)	1.015 (3.251)	0.0769 (3.374)	0.0198 (3.112)
<i>Idio.SUE</i> × <i>Passive</i>	2.314*** (0.252)	2.214*** (0.295)	2.275*** (0.323)	2.503*** (0.371)	2.714*** (0.382)
Observations	333,875	333,875	333,875	333,875	333,875
R-squared	0.034	0.035	0.035	0.035	0.035
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls	✓	✓	✓	✓	✓
Firm-Level Controls	✓	✓	✓	✓	✓
Weight	Value	Value	Value	Value	Value

Table 18 Sensitivity of earnings-response regressions to including a n -day post-earnings-announcement window. Estimates from:

$$100 \times r_{i,(t,t+n)} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} +$$

$$\gamma_1 (Sys.SUE_{i,t} \times Passive_{i,t}) \gamma_2 (Idio.SUE_{i,t} \times Passive_{i,t}) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}$$

where $r_{i,(t,t+n)}$ is the cumulative log market-adjusted return (in percentage points) from the effective earnings announcement date to $t + n$. $r_{i,(t,t+n)}$ is Winsorized at the 1% and 99% level by year. Controls in $X_{i,t}$ include 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 and firm fixed effect. Standard errors double clustered at the firm and year-quarter level in parenthesis.

E.3 Trends in pre-earnings turnover

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

$$AT_{i,t+\tau} = \alpha + \sum_{\tau=-21}^{22} \beta_{\tau} \mathbf{1}_{\{i,t+\tau\}} + e_{i,t+\tau} \quad (27)$$

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 15 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.

E.4 CRSP volume vs. total volume

A possible explanation for decreased pre-earnings turnover is that informed trading before earnings announcements has moved to dark pools. This could occur e.g., because on lit exchanges, informed traders are getting back-run by algorithmic traders (Weller, 2018). To test this, I obtained data on dark pool volume from FINRA. There does not appear to be an increase in dark pool volume in the weeks before earnings announcements, either in aggregate, or for stocks with high passive ownership.

E.5 IV estimates for mechanisms regressions

In Table 19, I reproduce all the regressions from the mechanisms section of the paper using the IVs built on Russell 1000 to 2000 switchers and S&P 500 index additions.

Panel A: Russell									
	IVD (1)	Bloomberg (2)	Downloads (3)	CAT (4)	Num Est (5)	SD Est (6)	Dist/SD(Est) (7)	Updates (8)	Time (9)
Passive Ownership	0.281*** (0.098)	-4.47*** (1.092)	13.02*** (2.879)	-16.09** (7.363)	-5.235 (4.552)	0.708* (0.399)	5.876*** (1.669)	0.253 (0.348)	-0.35 (1.459)
Panel B: S&P									
	IVD (1)	Bloomberg (2)	Downloads (3)	CAT (4)	Num Est (5)	SD Est (6)	Dist/SD(Est) (7)	Updates (8)	Time (9)
Passive Ownership	0.335** (0.139)	-1.1 (0.805)	24.57*** (2.133)	-14.72 (10.424)	111.5*** (11.220)	0.592*** (0.137)	6.561*** (1.792)	1.577*** (0.324)	-13.02*** (4.246)
Panel C: Earnings Responses									
	Russell (1)	S&P (2)							
<i>Idio.SUE</i> \times <i>Post</i> \times	18.22**	6.093**							
<i>Treated</i> \times <i>PassiveGap</i>	(8.565)	(2.508)							

Table 19 IV Estimates for mechanism regressions

For Panels A and B, estimates are from:

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

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

For Panel C, estimates are from:

$$100 \times r_{i,t} = \alpha + \beta_4 Post_{i,t} + \beta_5 Passive\ Gap_{i,t} \times Treated_{i,t} \times Post_{i,t} +$$

$$\beta_6 SUE_{i,t} + \beta_7 Passive\ Gap_{i,t} \times Treated_{i,t} \times Post_{i,t} \times SUE_{i,t} + FE + \epsilon_{i,t}$$

where $Post_{i,t}$ is an indicator for observations after the index change. $Passive\ Gap_{i,t}$ is the expected change in passive ownership from being treated. FE are fixed effects for each cohort. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

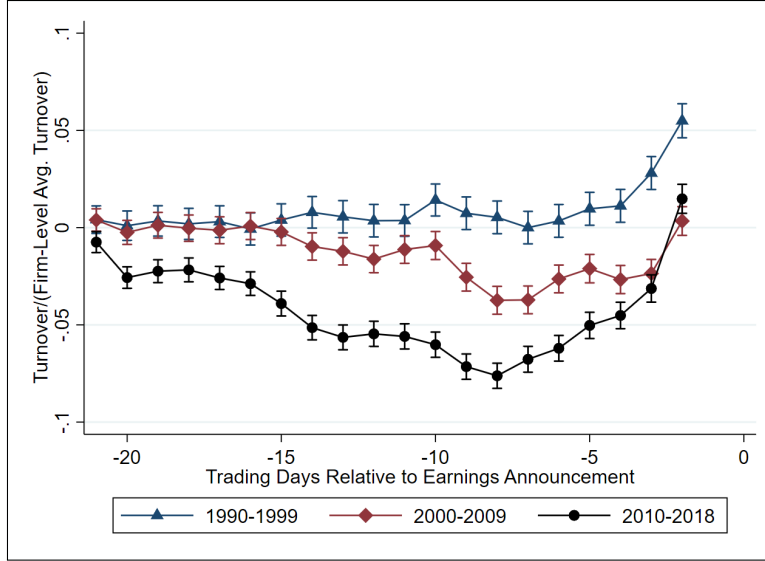


Figure 15. Decline of pre-earnings turnover. Plot of β_τ estimated from the regression:

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

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.

F *DM* and *QVS* in Kyle (1985)

In this section I show that, in the context of a 2-period Kyle (1985)-style model, *DM* and *QVS* are increasing in the precision of the insider's signal.⁴³

F.1 Model setup

The model has two trading periods, $t = 1$ and $t = 2$. There is a single risky asset whose value is distributed:

$$v \sim N(0, \sigma_v^2) \tag{28}$$

There is a *strategic* risk-neutral informed investor who receives an unbiased signal before

⁴³This section borrows heavily from Alex Chinco's "Two Period Kyle (1985) Model" notes.

the first trading period:

$$s = v + \epsilon \quad (29)$$

where v is the true value of the asset and ϵ is signal noise. ϵ is independent of v and normally distributed with mean zero and standard deviation σ_ϵ . This implies that $s \sim N(v, \sigma_\epsilon^2)$.

The informed investor submits demands to a set of *competitive* risk-neutral market makers 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 normally distributed with mean zero and standard deviation σ_z .

The set of competitive market makers observe total order flow x_t each period:

$$x_t = y_t + z_t \quad (30)$$

There is perfect competition among market makers, so they must set prices equal to the expected fundamental value of the asset given total demand:

$$p_1 = E[v|x_1] \quad \text{and} \quad p_2 = E[v|x_1, x_2] \quad (31)$$

In period 1, the informed investor chooses demand y_t to solve:

$$H_0 = \max_{y_1} E[(v - p_1) y_1 + H_1 | s] \quad (32)$$

where H_{t-1} is the informed investor's value function entering period t .

In period 2, they choose y_2 to maximize:

$$H_1 = \max_{y_2} E[(v - p_1) y_2 | s, p_1] \quad (33)$$

An equilibrium is made up of two components: (1) a linear demand rule for the informed investor in each period:

$$y_t = \alpha_{t-1} + \beta_{t-1} s \quad (34)$$

And (2) a linear pricing rule for the market makers in each period:

$$p_t = \kappa_{t-1} + \lambda_{t-1} x_t \quad (35)$$

The informed investor updates their beliefs about v after observing s . Their posterior beliefs about the mean and variance are:

$$\mu_{v|s} = \left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2} \right) \times \quad \text{and} \quad \sigma_{v|s}^2 = \left(\frac{\sigma_\epsilon^2}{\sigma_v^2 + \sigma_\epsilon^2} \right) \times \sigma_v^2 \quad (36)$$

where going forward, I will use θ in place of $\left(\frac{\sigma_v^2}{\sigma_v^2 + \sigma_\epsilon^2} \right)$.

The market makers extract an unbiased signal about v from total demand. Substituting in the informed trader's demand rule, the $t = 1$ signal is:

$$v = \frac{x_1}{\beta_0} - \epsilon - \frac{z_1}{\beta_0} \quad (37)$$

This implies that the market makers' posterior beliefs after observing x_1 are:

$$\mu_{v|x_1} = \left(\frac{\beta_0^2 \sigma_v^2}{\beta_0^2 \sigma_s^2 + \sigma_z^2} \right) \times x_1 \quad \text{and} \quad \sigma_{v|x_1}^2 = \left(\frac{\beta_0^2 \sigma_\epsilon^2 + \sigma_z^2}{\beta_0^2 \sigma_s^2 + \sigma_z^2} \right) \times \sigma_v^2 \quad (38)$$

Another way to think about this is that the total order flow x_1 is a signal about the informed trader's signal s rather than the fundamental value of the asset v . This would imply the $t = 1$ signal is:

$$s = \frac{x_1}{\beta_0} - \frac{z_1}{\beta_0} \quad (39)$$

which gives posterior beliefs:

$$\mu_{s|x_1} = \left(\frac{\beta_0^2 \sigma_s^2}{\beta_0^2 \sigma_s^2 + \sigma_z^2} \right) \times x_1 \quad \text{and} \quad \sigma_{s|x_1}^2 = \left(\frac{\sigma_z^2}{\beta_0^2 \sigma_s^2 + \sigma_z^2} \right) \times \sigma_s^2 \quad (40)$$

F.2 Solving the model

Given the market makers' zero profit condition, $\kappa_0 = 0$ and

$$\kappa_1 = E[v|x_1] - \lambda_1 E[x_2|x_1] = p_1 - (\theta \mu_{s|x_1} - p_1) = p_1 \quad (41)$$

where the last equality comes from $\theta \mu_{s|x_1} = p_1$.

Substituting in the market makers' linear pricing rule into H_1

$$H_1 = \max_{y_2} E[(v - \kappa_1 - \lambda_1 x_2) y_2 | s, p_1] \quad (42)$$

Taking the first order condition with respect to y_2 yields optimal demand:

$$y_2 = -\frac{p_1}{2\lambda_1} + \frac{\theta}{2\lambda_1}s \quad (43)$$

so $\alpha_1 = -\frac{p_1}{2\lambda_1}$ and $\beta_1 = \frac{\theta}{2\lambda_1}$.

With this, we can partially solve for the market makers' price impact coefficient, λ_t , in period 2:

$$\lambda_1 = \frac{Cov[x_2, v|x_1]}{Var[x_2|x_1]} = \frac{\beta_1 \sigma_{v|x_1}^2}{\beta_1^2 \sigma_{s|x_1}^2 + \sigma_z^2} \quad (44)$$

Now, turning to the period one solution, we start by taking a guess at the informed investors' value function which we will verify later:

$$E[H_1|s] = \phi_1 + \omega_1 (\mu_{v|s} - p_1)^2 \quad (45)$$

Substituting in the price impact and demand coefficients into H_0 yields:

$$H_0 = \max_{y_1} E[(v - p_1)y_1 + \phi_1 + \omega_1 (\theta s - p_1)^2 | s] \quad (46)$$

Taking the first order condition with respect to y_1 implies:

$$y_1 = \frac{\theta}{2\lambda_0} \left(\frac{1 - 2\omega_1\lambda_0}{1 - \omega_1\lambda_0} \right) s \quad (47)$$

With all this, we can now solve for the time 1 price impact coefficient:

$$\lambda_0 = \frac{Cov[x_1, v]}{Var[x_1]} = \frac{\beta_0 \sigma_v^2}{\beta_0^2 \sigma_s^2 + \sigma_z^2} \quad (48)$$

To verify the guess about H_1 , substitute the equilibrium coefficients for demands and prices into Equation 45:

$$H_1 = \left[\frac{1}{2\lambda_1} \left([v - \theta s] + \frac{1}{2}[\theta s - p_1] - \lambda_1 z_2 \right) (\theta s - p - 1) | s \right] \quad (49)$$

which simplifies to:

$$H_1 = \text{Constant} + \frac{1}{4\lambda_1} (\mu_{v|s} - p_1)^2 \quad (50)$$

This reveals that $\omega_1 = \frac{1}{4\lambda_1}$ and that H_1 is consistent with the original guess.

To solve the model, start with some initial guess for $\hat{\beta}_0$, and use this to compute the other equilibrium coefficients. This can be done in stages, first computing $\sigma_{v|x_1}^2$ and $\sigma_{s|x_1}^2$, and then using these to compute $\hat{\lambda}_1$:

$$\begin{aligned}\hat{\lambda}_0 &= \frac{\hat{\beta}_0 \sigma_v^2}{\hat{\beta}_0^2 \sigma_s^2 + \sigma_z^2} \\ \sigma_{v|x_1}^2 &= \frac{\hat{\beta}_0^2 \sigma_\epsilon^2 + \sigma_z^2}{\hat{\beta}_0^2 \sigma_s^2 + \sigma_z^2} \sigma_v^2 \\ \sigma_{s|x_1}^2 &= \frac{\sigma_z^2}{\hat{\beta}_0^2 \sigma_s^2 + \sigma_z^2} \sigma_s^2 \\ \hat{\lambda}_1 &= \frac{1}{\sigma_z} \sqrt{\frac{\theta}{2} \left(\sigma_{v|x_1}^2 - \frac{\theta}{2} \sigma_{s|x_1}^2 \right)}\end{aligned}\tag{51}$$

A solution has been found when you have minimized the distance between the guess $\hat{\beta}_0$ and $\frac{\theta}{2\hat{\lambda}_0} \left(\frac{1-2\omega_1\hat{\lambda}_0}{1-\omega_1\hat{\lambda}_0} \right)$, which is a condition $\hat{\beta}_0$ has to satisfy in equilibrium.

F.3 Mapping the model to DM and QVS

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 $p_3 = v$ i.e., the terminal payoff. 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 periods. I view $t = 3$ as the date where the fundamental value of the firm is revealed i.e., an earnings announcement.

Pre-Earnings Drift

In this setting, a natural definition of the pre-earnings drift would be

$$DM = \begin{cases} \frac{1+r_{(1,2)}}{1+r_{(1,3)}} & \text{if } r_{(2,3)} > 0 \\ \frac{1+r_{(1,3)}}{1+r_{(1,2)}} & \text{if } r_{(2,3)} < 0 \end{cases} \Leftrightarrow \begin{cases} \frac{1}{1+r_{(2,3)}} & \text{if } r_{(2,3)} > 0 \\ 1 + r_{(2,3)} & \text{if } r_{(2,3)} < 0 \end{cases}\tag{52}$$

In this model, both the informed investor and market makers are risk neutral, so prices can be negative. With this in mind, I define the return from period a to b as $r_{(a,b)} = p_b - p_a$. The definition in Equation 52, however, is based on percentage returns. To account for this

difference, I leverage the fact that DM only depends on the magnitude of $r_{(2,3)}$ (ignoring the slight asymmetry with respect to positive and negative returns), and re-define $DM = 1 - |r_{(2,3)}|$.

DM should be related to the precision of the informed investor's signal because of the market makers' 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 .

Share of Volatility on Earnings Days

The natural mapping of QVS to this setting is:

$$QVS = 1 - \frac{r_{(2,3)}^2}{r_{(1,2)}^2 + r_{(2,3)}^2} \quad (53)$$

Again, however, given that returns are defined as level changes in prices, rather than percentage changes in prices, I re-define $QVS = 1 - \frac{|r_{(2,3)}|}{|r_{(1,2)}| + |r_{(2,3)}|}$. In words, QVS is the fraction of the total distance traveled between $t = 1$ and $t = 3$ that occurs before uncertainty is resolved. Consistent with the intuition outlined above, as the insider's signal becomes 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 .

F.4 Simulation results

For each set of parameters, I simulate the economy 10,000 times and compute averages of DM and QVS . The left panel of Figure 16 shows the relationship between DM and the volatility of the informed investor's signal noise, σ_ϵ . The right panel is similar, plotting the relationship between QVS and σ_ϵ . Consistent with the intuition outlined above, DM and QVS are monotonically decreasing in σ_ϵ . Although this is just one set of parameters, I find these relationships are monotonic across a broad set of possible parameter choices.

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

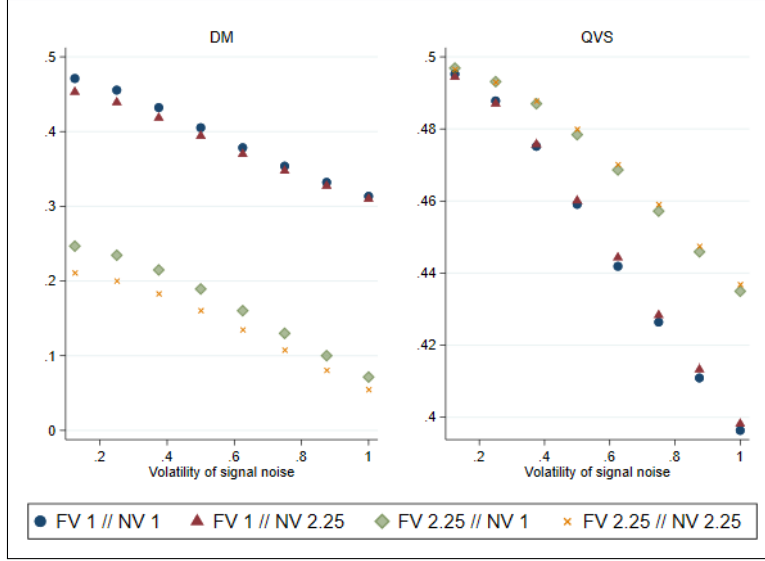


Figure 16. Informed investors' precision, DM and QVS . Each point represents the average of the price informativeness measures across 10,000 simulations. FV =fundamental volatility= σ_v^2 . NV =noise volatility= σ_z^2 .

because more 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 both panels of Figure 16.

Differences between DM and QVS

Figure 16 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.⁴⁴

⁴⁴This 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 $z_1 > 0$, as well as a draw of v close to \bar{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.