Passive Ownership and Price Informativeness

MARCO SAMMON *

October, 2023

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 approximately $1/4^{th}$ of its whole sample mean and $1/6^{th}$ of its whole sample standard deviation.

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

*Harvard Business School. I would like to thank my dissertation committee: Scott Baker, Alex Chinco, Ravi Jagannathan, Robert Korajczyk and Dimitris Papanikolaou. I would also like to thank 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 2019, 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. This approach is motivated by early studies of market efficiency (Ball and Brown, 1968), leveraging the fact that earnings announcements are times where we know large quantities of information are released to the market. I quantify the amount of information incorporated into prices prior to earnings announcements in four ways, which range from a precise measure of how surprising the earnings news was to markets, to measures which are noisier, but closer to theories of price informativeness (Grossman and Stiglitz (1980), Kyle (1985)).

First, I examine the absolute market-adjusted earnings day return, $|Ret|$. This is a natural measure of how surprised markets were by earnings news, with the logic being that if less information was incorporated into prices ahead of time, the earnings day return should be larger (Frazzini, 2006). $|Ret|$ is a precise measure of the surprise, because earnings news is likely the main driver of returns on earnings-announcement days.

As a measure of pre-earnings announcement price informativeness, $|Ret|$ has several disadvantages. One is that $|Ret|$ is sensitive to both cross-sectional and time series variation in volatility. To this end, I also examine the absolute magnitude of the earnings day return divided by the standard deviation of returns over the 22 trading-days before the announcement itself, $|Ret|/SD$. This quantity may be a slightly less precise measure of how surprising the earnings news was than $|Ret|$, because factors other than the incorporation of earnings news may affect volatility over the month before the earnings announcement. But, $|Ret|/SD$ accounts for differences in average volatility across stocks, and time-series variation in aggregate volatility.

Normalizing the earnings-day return by the standard deviation of past returns, however, still doesn't account for differences in cumulative pre-earnings returns. Consider, for example, two stocks with the same pre-earnings volatility, and the same earnings day return of 10%. Suppose one of these stocks had a pre-earnings cumulative return of 0%, while the other had a return of 10% over the same period. The stock with the larger earnings day return *relative to the pre-earnings return* (i.e., the stock with a 0% pre-earnings return) intuitively seems to have had a less informative pre-earnings announcement price.

To account for differences in pre-earnings returns, I compute the fraction of the total net return from a month before the earnings announcement to two days after the earnings announcement, which occurs after the earnings release. This is the price-jump measure of Weller (2018), PJ. One interpretation of PJ is the share of earnings information which was incorporated into prices after the information was made public.

PJ is a more ambitious measure than $|Ret|$ and $|Ret|/SD$, which are only measuring how surprising the earnings news was to markets. By measuring the fraction of the total information left in prices before the earnings was formally released, PJ is closer to measures of price informativeness that arise in Grossman and Stiglitz (1980) and Kyle (1985)-style models, which speak to the conditional volatility of future fundamentals given current prices. That being said, PJ is likely noisier than e.g., $|Ret|$ as many factors other than the incorporation of earnings information could drive pre-earnings returns.

All the three measures so far leverage realized earnings-day returns to evaluate pre-earnings announcement price informativeness. Options markets, however, can be used to evaluate investors' ex-ante beliefs about earnings announcements. If fewer investors are gathering private information ahead of earnings announcements, one would expect more ex-ante earnings uncertainty and the implied volatility of options which span earnings announcements should be relatively higher (Dubinsky et al., 2006). To this end, I adapt the Implied

Volatility Difference measure of Kelly et al. (2016), IVD, to evaluate whether options which span earnings announcements are more expensive than those which expire before and after the announcement. Like PJ, IVD is sensitive to many other factors that could affect the prices of options which expire more than a month before or after earnings announcements themselves. That being said, by subtracting out the average level of implied volatility both before and after the announcement, IVD should account for time series and cross-sectional differences in volatility.

In summary, $|Ret|$, $|Ret|/SD$, PJ and IVD are the key measures I use to evaluate the effect of passive ownership on pre-earnings announcement price informativeness. I interpret higher values of all these measures as evidence that less private information was incorporated into prices 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 2019. 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, average $|Ret|$ was roughly 2% while by 2019 this average had grown to 4%, an increase of roughly $1/3^{rd}$ of $|Ret|$'s whole sample standard deviation. This pattern also holds for average $|Ret|/SD$, which increased from just over 1 in 1990 to 3.5 by 2019. This increase was on par with $|Ret|/SD$'s whole sample standard deviation of 2.6. PJ also increased on average over my sample, going from 0.2 to 0.5, an increase about equal to PJ's whole sample standard deviation. Finally, average IVD increased from 0.02 to 0.05, about $1/3^{rd}$ of IVD's whole sample standard deviation. All four measures indicate that on average, less information is being incorporated into prices ahead of earnings announcements in recent years, relative to the early 1990s.

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 2019 has a 1.06 higher $|Ret|/SD$ than a firm in the 10th percentile of passive ownership in 2019. For reference, the difference in passive ownership share between these two percentiles is 24%. This effect is large, at more than half of $|Ret|/SD$'s whole sample mean of 2.00 and over 40% of its whole sample standard deviation of 2.6. Once again, the results for $|Ret|$, PJ and IVD corroborate these findings.

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. 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, 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. In fact, across all four measures, the IV estimates are 2.5-4 times as large as the OLS estimates. 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.

Of course, index switching may be correlated with many other factors which could affect pre-earnings announcement price informativeness. For example, being added to the S&P 500 could be correlated with increased analyst coverage, investor attention or changes in disclosure. While all these things are true for firms added to the S&P 500, the opposite is true for firms which switch from the Russell 1000 to the Russell 2000. One reason for this is that firms added to the S&P 500 are growing, while those added to the Russell 2000 are shrinking. Despite these differences, the IV estimates from both S&P index changes and Russell index changes are similar in magnitude, suggesting that a seeming violation of the only-through assumption is likely not driving my results.

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.[1] 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 estimated effects across the four measures of price informativeness, a 15% increase in passive ownership (i.e., the value weighted increase in passive ownership over my sample) has caused the amount of information incorporated into prices ahead of earnings announcements to decline by approximately $1/4^{th}$ of its whole sample mean and $1/6^{th}$ of its whole sample standard deviation.

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

[1] For 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.

and Pedersen (2018), Kacperczyk et al. (2018b), Breugem and Buss (2019), Brogaard et al. (2019), Bennett et al. (2020a), Bennett et al. (2020b), Kothari et al. (2023)) or non-existent link (Coles et al., 2022).[2] 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 contribution of my paper is to use earnings announcements as a laboratory to study not just the effect of passive ownership on price informativeness, but also how passive ownership affects *when* information is incorporated into prices. To this end, my measures of price informativeness focus specifically on the narrow window ahead of earnings announcements and quantify 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.

2 Measurement & data

This section motivates the four 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 2019.

2.1 Measurement

A natural way to measure how surprising earnings news was to the market is to look at the absolute magnitude of abnormal earnings-day returns (Frazzini, 2006). Specifically, for

[2] The two most closely-related papers are Israeli et al. (2017) and Glosten et al. (2021), as both are studying the effects of passive ownership on how stocks respond to earnings news. Specifically, Israeli et al. (2017) and Glosten et al. (2021) are interested in the contemporaneous relationship between passive ownership and earnings responses. In other words, these papers are focused on the way stock prices respond *conditional* on a given amount of information being released. My paper differs from both of these papers, in that I am focused on the share of information incorporated into prices *ahead* of earnings announcements.

each firm i on earnings announcement day t, I compute the absolute market-adjusted return, $|Ret|_{i,t}$, as the difference between the return on stock i and the return on the CRSP value-weighted index (Campbell et al., 2001). The assumption underlying this interpretation of $|Ret|_{i,t}$ is that earnings announcement news is fully incorporated into prices quickly after its release (within a day). In this case, the bigger the earnings-day return, the larger the amount of information which was not incorporated into prices ahead of time. This interpretation of $|Ret|_{i,t}$ 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 informativeness, it should cause* $|Ret|_{i,t}$ *to increase*

Using $|Ret|_{i,t}$ to study pre-earnings announcement price informativeness has two main benefits. First, because it focuses only on the earnings-day return, it is unlikely to be contaminated by the stock's response to other types of information. Second, because it is straightforward to observe and only uses the announcement-day return, it can be computed for every earnings announcement.

As a measure of pre-earnings announcement price informativeness, $|Ret|_{i,t}$ also has several limitations. First, $|Ret|_{i,t}$ 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 ± 1% while stock B has alternating returns of ± 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 $|Ret|_{i,t}$. This is an especially important concern in my setting, as past research has shown that passive ownership affects stock-level volatility (Ben-David et al., 2018).

To address this limitation, I also examine the earnings day return divided by the standard deviation of returns over the 22 trading days before the earnings announcement itself:

$$|Ret|/SD_{i,t} = |r_{i,t}|/\sigma_{(t-22,t-1)}\left(r_{i,t}\right) \tag{1}$$

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

Like $|Ret|_{i,t}$, 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 relative to pre-announcement volatility, and therefore $|Ret|/SD_{i,t}$ should be smaller as well. This interpretation of $|Ret|/SD_{i,t}$ yields the second empirical prediction I use to measure the relationship between passive ownership and price informativeness.

Prediction 2: *If passive ownership decreases pre-earnings announcement price informativeness, it should cause $|Ret|/SD_{i,t}$ to increase*

$|Ret|/SD_{i,t}$ is likely a less precise measure of pre-earnings announcement price informativeness than $|Ret|_{i,t}$ because factors other than the incorporation of earnings news might drive pre-earnings announcement volatility. That being said, using $|Ret|/SD_{i,t}$ has several advantages. First, like $|Ret|_{i,t}$, it is easily observable and can be computed for every earnings announcement with non-missing prior return data. Second, it accounts not only for cross-sectional heterogeneity in volatility, but for time-series variation in volatility as well. Specifically, during periods of high volatility, $|Ret|_{i,t}$ might be higher for all stocks. But, if this increase is proportional to the unconditional increase in volatility, $|Ret|/SD_{i,t}$ should not be affected.

Using $|Ret|/SD_{i,t}$ also has limitations, with one potential downside being that it does not condition on the cumulative pre-earnings announcement return (an issue shared by $|Ret|_{i,t}$). For example, consider two stocks that have a 10% earnings-day return. Stock A has a cumulative return of 0% over the 22 trading days before the announcement, while stock B has a cumulative return of 10% over the same period. If both these stocks have the same pre-announcement volatility, they would have the same $|Ret|/SD_{i,t}$. From the perspective of measuring price informativeness, however, this seems counterintuitive given that stock B's earnings-day return is smaller than stock A's, *relative to its pre-earnings announcement run-up*. And, these large pre-announcement returns could have been due to earnings information being incorporated into prices before it was formally released.

Accounting for pre-announcement returns speaks to a broader literature which has shown that prices incorporate a substantial portion of earnings news before it is actually made public (Ball and Brown, 1968). A natural measure of pre-earnings announcement price informativeness, therefore, is the percentage of the total information which was incorporated into prices after the news was released. This intuition motivates the price jump measure of

Weller (2018):

$$PJ_{i,t} = \frac{CAR_{i,t}^{(T-1,T+b)}}{CAR_{i,t}^{(T-a,T+b)}} \tag{2}$$

where $CAR_{i,t}^{(k_1,k_2)}$ is the cumulative abnormal return from dates k_1 to k_2 around announcement date T. In words $PJ_{i,t}$ is the fraction of the total *net* return from $T-a$ to $T+b$ that occurs *after* the earnings announcement, with higher values implying that more of the information was incorporated into prices after the information was made public. This interpretation of $PJ_{i,t}$ yields the third empirical prediction I use to measure the effect of passive ownership on price informativeness.

Prediction 3: *If passive ownership decreases pre-earnings announcement price informativeness, it should cause $PJ_{i,t}$ to increase*

One advantage of $PJ_{i,t}$ relative to $|Ret|_{i,t}$ and $|Ret|/SD_{i,t}$ is that it speaks to a more ambitious question: what *share* of the earnings information was incorporated into prices ahead of time. $|Ret|_{i,t}$ and $|Ret|/SD_{i,t}$, on the other hand, only speak to how surprising the news was to markets. Intuitively, PJ seems closer to the measures of price informativeness that arise in a Grossman and Stiglitz (1980) or Kyle (1985)-style model, which are designed to capture the conditional volatility of fundamentals given prices before uncertainty is resolved. That being said, $PJ_{i,t}$ is likely a noisier measure of pre-earnings announcement price informativeness than $|Ret|_{i,t}$, as many non-earnings-related factors can affect realized returns in the month ahead of an earnings announcement (e.g., macro news announcements, FOMC meetings and other types of company-specific news).

Another issue with $PJ_{i,t}$ is that it is not well defined when $CAR_{i,t}^{(T-a,T+b)}$ is close to zero. Weller (2018) solves this by removing "non-events" i.e., earnings announcements where the total cumulative return from $T-a$ to $T+b$ is near zero. This filter, however, removes the majority of earnings announcements in his sample (54.5%).

Another possible drawback of $PJ_{i,t}$ is that it is sensitive to the level of volatility in a different way than $|Ret|/SD_{i,t}$, because it is normalizing by the total return, rather than the volatility of returns. For example, 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, stocks A and B both have a return of 1%. These stocks will have the same $PJ_{i,t}$ of 1, even though it seems like the 1% announcement day return is more surprising for the stock which had less pre-announcement volatility. This, however, will be picked up by $|Ret|/SD$, which is $1/5^{th}$ as large for the more volatile stock.

As mentioned above, $PJ_{i,t}$ may be affected by noise in realized pre-announcement returns. $|Ret|_{i,t}$ accounts for this noise by focusing on just one day where the earnings announcement news is likely the main driver of stock price changes. A different way to account for noise in realized returns is to use options data. Options have the advantage of being priced based on the market's ex-ante expectation of volatility, rather than ex-post realizations of volatility/returns. If there is more ex-ante earnings uncertainty, we would expect options exposed to earnings announcement risk to be relatively more expensive (Dubinsky et al., 2006). To quantify this, 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. Then, the implied volatility difference is defined as:

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right) \tag{3}$$

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.[3]

My preferred interpretation of IVD is built on the same logic as $|Ret|_{i,t}$ and $|Ret|/SD_{i,t}$. 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 $IVD_{i,t}$ and passive ownership.

[3]See Appendix A.4 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 all my regression specifications. In addition, all the results are qualitatively unchanged instead defining the implied volatility difference as a ratio: $\widetilde{IVD}_{i,\tau} = \overline{IV}_{i,b}/\frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right)$.

Prediction 4: *If passive ownership decreases pre-earnings announcement information gathering, it should cause $IVD_{i,t}$ to increase*

As mentioned above, an advantage of $IVD_{i,t}$ is that it does not rely on realized returns on the earnings-day itself, which could be contaminated by contemporaneous information releases. Another advantage of $IVD_{i,t}$ is that, because it subtracts the average implied volatility of options that expire before and after the earnings announcement, it should account for cross-sectional and times-series variation in volatility as well.

That being said, $IVD_{i,t}$ also has several limitations. One issue is that $IVD_{i,t}$ can only be computed for a relatively small subset of earnings announcements. The original method in Kelly et al. (2016) was designed for index options, which typically have much richer coverage than options on individual equities, so the sample of earnings announcements with sufficient coverage to compute $IVD_{i,t}$ is less than half of the overall sample of earnings announcements I consider. In addition, because $IVD_{i,t}$ uses data from such a large range of dates around earnings announcements, it is also likely noisy, in the sense that it could be confounded by the options spanning other non-earnings news events.

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 2019.[4] For each stock, around each earnings announcement, I need to construct $|Ret|$, $|Ret|/SD$, PJ, IVD 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, along with

[4] I discuss the effect of adding 2020-2022 to my sample in Appendix E.

the return on the CRSP value-wighted index, to construct $|Ret|$, $|Ret|/SD$ and PJ.[5] All options data is from OptionMetrics. The sample for IVD is shorter than for the other three measures of pre-earnings announcement price informativeness because OptionMetrics coverage begins in 1996.

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 Appendix C.3, I show that my results are quantitatively unchanged by restricting only to funds with an index fund flag of "D" in the CRSP mutual fund database (Crane and Crotty, 2018).

2.3 Basic properties

To visualize the time-series and cross-sectional properties of the five key variables in my analysis, Figure 1 plots the 25th percentile, median, 75th percentile and value-weighted average of $|Ret|$, $|Ret|/SD$, PJ, IVD and the level of passive ownership. The top left panel shows that passive ownership steadily increased over my sample. From 1990 to 2019, 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. 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 2019.

The top middle panel shows that average $|Ret|$ increased by about 0.02 between 1990 and 2019. This increase is roughly $1/3^{rd}$ of $|Ret|$'s whole sample standard deviation of 0.062. There are notable spikes in average $|Ret|$ in the early 2000s and again in the late 2000s. One explanation for these spikes is that these years correspond to the dot-com boom and the

[5] I modify Weller (2018)'s original implementation of PJ in two ways: (1) to avoid sensitivity to estimating betas, I use market-adjusted returns instead of factor-model adjusted returns (2) for consistency with my other measures of pre-earnings announcement price informativeness, I use $a = 22$ instead of $a = 21$ (like the original paper, I set $b = 2$). I also, therefore, apply the non-event filter using past volatility as of $T - 22$ instead of $T - 21$.

Global Financial Crisis. These were periods with higher overall volatility, leading to larger absolute earnings-day returns on average.

The top right panel shows that average $|Ret|/SD$ increased from just above 1 in 1990 to about 3.5 in 2019. This rise is of similar magnitude to $|Ret|/SD$'s whole sample standard deviation of 2.6. Unlike $|Ret|$, $|Ret|/SD$ does not have a huge spike in the dot-com boom, possibly because dividing by the standard deviation of pre-announcement returns explicitly accounts for the level shift up in realized volatility. As with passive ownership, there has also been a trend toward increased cross-sectional spread in $|Ret|/SD$, with the interquartile range increasing from 1.215 in the 1990s to 3.258 in the 2010s. The time series increase in $|Ret|/SD$ 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).

The bottom left panel of Figure 1 shows that, consistent with a trend toward decreased pre-earnings announcement price informativeness, average PJ has increased over my sample from 0.2 to over 0.5. This is an economically large increase, nearly the same magnitude as PJ's whole sample standard deviation of 0.479. As with $|Ret|/SD$, the cross-sectional spread in PJ has steadily increased over the past 30 years.

Finally, the bottom middle panel shows that IVD increased over my sample. From 1990 to 2019, average IVD went from about 0.02 to 0.05, close to $1/3^{rd}$ if IVD's whole sample standard deviation of 0.113. Like $|Ret|$, IVD spikes during the Great Financial Crisis, and during the dot-com boom/bust.

The findings in Figure 1 seem striking, as previous literature (e.g., Bai et al. (2016), Dávila and Parlatore (2018)) have shown a time series trend toward increased price informativeness. 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

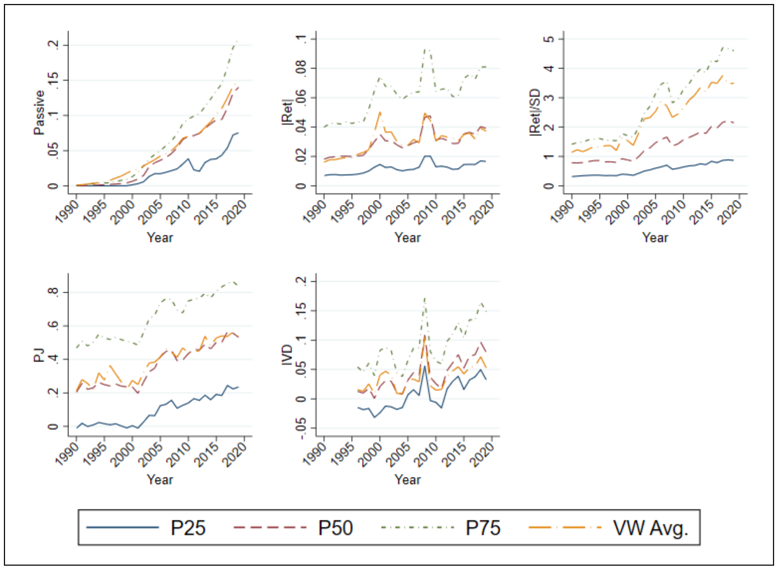

Figure 1. Trends in $|Ret|$, $|Ret|/SD$, PJ, IVD and the level of passive ownership, 1990-2019. To compute the value-weighted average (*VW Avg.*), within each quarter, observations are weighted in proportion to their market capitalization at the end of the previous quarter. Passive ownership is defined 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. $|Ret|$ is the absolute earnings-day return. $|Ret|/SD$ is the absolute earnings-day return divided by the standard deviation of returns over the 22 trading days before the earnings announcement. $PJ_{i,t}$ is Weller (2018)'s price jump measure, defined in Equation 2. $IVD_{i,t}$ is Kelly et al. (2016)'s implied volatility difference measure, defined in Equation 3.

information being incorporated into prices after the news is released.[6]

Table 1 contains summary statistics on the measures of pre-earnings announcement price informativeness, as well as passive ownership. As mentioned above, the sample with non-

[6]My results are not necessarily 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 $|Ret|$, $|Ret|/SD$ and PJ 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 Parlatore (2018).

missing PJ is less than half the overall sample, because of the non-event filter. The sample for IVD is even smaller, owing both to having fewer years of data, and having a limited number of equities with sufficiently rich option coverage. Appendix B.1 contains details on the correlations between the measures of pre-earnings announcement price informativeness.

		# Observations	p25	p50	mean	p75	sd
1990-1999	Passive	155,670	0.000	0.001	0.003	0.004	0.004
	$\|Ret\|$	155,670	0.008	0.021	0.036	0.046	0.049
	$\|Ret\|/SD$	155,670	0.352	0.827	1.201	1.567	1.388
	PJ	44,056	0.007	0.241	0.270	0.514	0.423
	IVD	16,991	-0.022	0.008	0.019	0.049	0.089
2000-2009	Passive	162,373	0.009	0.026	0.033	0.049	0.030
	$\|Ret\|$	162,373	0.013	0.032	0.054	0.071	0.070
	$\|Ret\|/SD$	162,373	0.502	1.205	2.007	2.550	2.475
	PJ	56,579	0.070	0.347	0.374	0.663	0.479
	IVD	43,067	-0.005	0.034	0.049	0.089	0.108
2010-2019	Passive	123,195	0.041	0.088	0.093	0.133	0.067
	$\|Ret\|$	123,195	0.014	0.033	0.053	0.070	0.064
	$\|Ret\|/SD$	123,195	0.755	1.880	2.998	4.040	3.394
	PJ	49,478	0.184	0.493	0.500	0.806	0.500
	IVD	52,026	0.022	0.061	0.076	0.119	0.121
Full Sample	Passive	441,238	0.002	0.014	0.039	0.060	0.054
	$\|Ret\|$	441,238	0.011	0.028	0.048	0.061	0.062
	$\|Ret\|/SD$	441,238	0.484	1.155	1.999	2.469	2.579
	PJ	150,113	0.075	0.357	0.385	0.677	0.479
	IVD	112,084	0.002	0.043	0.057	0.100	0.113

Table 1 Summary Statistics. Cross-sectional equal-weighted means, standard deviations and distributions of pre-earnings announcement price informativeness measures 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 $|Ret|$,

$|Ret|/SD$, PJ and IVD on passive ownership. Across all four measures, the regressions show that higher passive ownership is correlated with decreased pre-earnings announcement price informativeness. I then discuss robustness checks showing 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:

$$\text{Price informativeness}_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t} \tag{4}$$

where Price informativeness$_{i,t}$ is either $|Ret|$, $|Ret|/SD$, PJ or IVD. $Passive_{i,t}$ is passive ownership from the last calendar month end before the earnings announcement itself. 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. The controls in $X_{i,t}$ are selected to capture firm characteristics known to be correlated with passive ownership (Glosten et al., 2021). Appendix A.1 contains details on the construction of all control variables, while Appendix B.2 documents the correlations between the stock-level controls and passive ownership.

Equation 4 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 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 1, column 1 shows that there is a positive relationship between passive ownership and $|Ret|$. The point estimate implies that a firm in the 90th percentile of passive ownership in 2019 (27%) has a 152 basis point higher $|Ret|$ than a firm in the 10th percentile of passive ownership in 2019 (3%). For reference, 152 basis points is roughly $1/3^{rd}$ of $|Ret|$'s whole sample mean of 476 basis points and $1/4^{th}$ of its whole sample standard deviation of 621 basis points. To allay concerns that small firms are driving my results, Column 5 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.[7]

Column 2 shows that, consistent with prediction 2, there is also a positive correlation between $|Ret|/SD$ and passive ownership. The point estimate implies that a firm in the 90th percentile of passive ownership in 2019 has 1.062 higher $|Ret|/SD$ than a firm in the 10th percentile of passive ownership in 2019. For reference, 1.062 is approximately 1/2 of $|Ret|/SDs$'s whole sample mean of 1.999 and $2/5^{ths}$ of its whole sample standard deviation of 2.579. Column 6 shows that the relationship between passive ownership and $|Ret|/SD$ is robust to using value weights, instead of equal weights.

Consistent with prediction 3, column 3 shows a positive relationship between passive ownership and PJ. In terms of magnitudes, a firm in the 90th percentile of passive ownership in 2019 has 0.0595 higher PJ than a firm in the 10th percentile of passive ownership in 2019. For reference, 0.0595 is approximately $1/6^{th}$ of PJ's whole sample mean of 0.395 and $1/8^{th}$ of its whole sample deviation of 0.479. I do not report a value-weighted version of Column 3, as the non-event filter is negatively correlated with firm size (i.e., large firms are more likely to have "non-event" earnings announcements), and can lead the remaining large firms to have within-quarter weights of over 10%.

Finally, column 4 shows that, consistent with prediction 4, IVD is positively correlated with passive ownership. The estimated coefficient implies that a firm in the 90th percentile of passive ownership in 2019 has a 0.0268 higher average IVD than a firm in the 10th percentile of passive ownership in 2019. This is slightly less than half of IVD's whole sample mean of 0.057 and about a $1/4^{th}$ of its whole sample standard deviation of 0.113. Using value weights instead of equal weights increases the coefficient on $Passive$ to 0.179. The value-weighted mean and standard deviation of IVD are 0.0366 and 0.0642, so the effect of passive ownership on IVD is relatively larger on a value-weighted basis.

As shown in Figure 1, value weighted average passive ownership increased by 15% between 1990 and 2019. Based on the OLS regressions in Table 2, a 15% higher level of passive

[7]It is important to highlight that the cross-sectional means and standard deviations are also different on a value-weighted basis. For example, although the point estimate in column 5 is smaller than in column 1, the value-weighted mean of $|Ret|$ is 306 basis points, and the value-weighted standard deviation of $|Ret|$ is 353 basis points, both of which are smaller than their equal weighted counterparts.

	$\|Ret\|$ (1)	$\|Ret\|/SD$ (2)	PJ (3)	IVD (4)	$\|Ret\|$ (5)	$\|Ret\|/SD$ (6)
Passive	0.0635***	4.425***	0.248***	0.112***	0.0572***	2.717**
	(0.007)	(0.384)	(0.060)	(0.030)	(0.013)	(1.281)
Observations	441,238	441,238	148,864	111,604	441,238	441,238
R-squared	0.194	0.214	0.163	0.3	0.27	0.227
Weight	Equal	Equal	Equal	Equal	Value	Value
Controls	YES	YES	YES	YES	YES	YES
Fixed Effects	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ

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

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

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$. 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.

ownership is correlated with a 95 basis point higher $|Ret|$, a 0.66 higher $|Ret|/SD$, a 0.037 higher PJ and a 0.0167 higher IVD. For $|Ret|$, this is 0.2 of its whole sample mean of 476 basis points and 0.15 of its whole sample standard deviation of 621 basis points. For $|Ret|/SD$, this is 0.33 of its whole sample mean of 2.0 and 0.26 of its whole sample standard deviation of 2.6. For PJ, this is 0.10 of its whole sample mean of 0.39 and 0.08 of its whole sample standard deviation of 0.48. For IVD, this is 0.29 of its whole sample mean of 0.06 and 0.15 of its whole sample standard deviation of 0.11. An equal weighted average of these effects relative to variables' means is 0.23 (roughly $1/4^{th}$), while an equal weighted average of these effects relative to variables' standard deviations is 0.16 (roughly $1/6^{th}$).

So far, I have only discussed the magnitudes of the estimated effects relative to those variables' own means and standard deviations. Given that the literature on the relationship between pre-earnings announcement price informativeness has come to so many different conclusions, it is also useful to compare my estimates with those in the literature. This, however, is not straightforward, as each paper measures price informativeness in a different way, and most of the papers are not focused on pre-earnings announcement price informativeness.

One of the exceptions to this is Glosten et al. (2021), who study the effect of passive ownership on earnings responses. The authors find that going from bottom decile to top decile of change in ETF ownership increases earnings response by 0.11. This is relative to an average response of 0.240. This is of similar magnitude to my estimate that going from the bottom to top decile of passive ownership decreases price informativeness by about 50% of its mean.

Another natural point of comparison is Bennett et al. (2020a), who study the effect of being added to the S&P 500 on price informativeness. Using the authors' estimates, a 24% increase in passive ownership predicts a decline in GPIN (Duarte et al., 2020) of about 0.09, 30% higher than GPIN's mean of 0.07 and 50% higher than GPIN's standard deviation of 0.17. So, these effects are larger than my estimated effects of passive ownership on $|Ret|/SD$, but are within the same order of magnitude.

Finally, there are papers which document no relationship between passive ownership and price informativeness (Coles et al., 2022), or a negative relationship between passive ownership and price informativeness (e.g., 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), Kothari et al. (2023)). So, relative to these papers, my estimated effects are large, but I would like to highlight that, with one exception, none of these papers are focused on the effect of passive ownership on pre-earnings announcement price informativeness.[8]

3.2 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

[8]The exception is Kacperczyk et al. (2018b), who find that all institutional ownership increases price informativeness, although the effect is stronger for active investors than passive investors. And, Kacperczyk et al. (2018b) are using also using the price jump measure from Weller (2018). The authors find that the MSCI shock predicts a decrease in PJ of 0.041 and the JGTRRA shock predicts a decrease in 0.225. This is relative to PJ's mean of 0.39 and standard deviation of 0.55 (which are similar to the mean and standard deviation of PJ in my sample). Their experiments, however, are contaminated by multiple effects because the MSCI shock translates to a 0.013 increase in active ownership and a 0.006 increase in passive ownership, while the JGTRRA shock is associated with an 0.010 increase in active ownership and 0.0004 increase in passive ownership. So price informativeness could be increasing because these shocks lead to increases in active ownership, which in both cases are significantly larger than the associated increase in passive ownership.

have differently affected stocks with more passive ownership. In Appendix C.2.1, columns 9 to 12 of Table C1 show that the OLS estimates are qualitatively unchanged when using only earnings announcements between 2001 and 2019, 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 – especially when using PJ to measure pre-earnings announcement price informativeness – if e.g., high passive stocks also have high AT activity due to ETF arbitrage. In Appendix C.2.2, columns 1 to 8 replicate the baseline regressions, but explicitly control for the AT measures in Weller (2018).[9] The OLS estimates are not significantly changed by including these controls, 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.7 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) (Chinco and Fos, 2021). Ex-ante, it's not obvious why the intersection of these rules would select stocks with low pre-earnings announcement price informativeness.

[9] These 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.

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 increases the fraction of a firm's shares that need to be held by passive funds. This is because the Russell 2000 has a higher average passive ownership share than the Russell 1000 (Pavlova and Sikorskaya, 2022). The logic is that a value weighted index fund holds the same fraction of each of the index constituent's shares outstanding (ignoring float adjustments). So, if one index has a higher level of passive ownership than another, switching to that index will tend to increase passive ownership.

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 use data on index membership from FTSE Russell and follow the method in Coles et al.

(2022).[10] Using their May market capitalization proxy, I correctly predict Russell 1000/2000 index membership for 99.24% of 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 2007-2019, 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.[11] 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.[12] These filters yield 724 treated firms and 618 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.[13]

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. This makes using additions to the S&P 500 a less clean laboratory than firms switching between the Russell 1000 and 2000, as it is not based on a purely mechanical rule. And if, for example,

[10] I would like to thank the authors for sharing their replication code with me. Appendix D.2 contains a step-by-step explanation of how I compute the May market capitalization proxy.

[11] Another 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 Appendix D.3.1, I show that within one year of switching, this decrease is totally offset by the time trend toward increased passive ownership.

[12] One 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.

[13] A 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.

the committee selects firms based on fundamental volatility, or another quantity related to price informativeness, the only-through assumption underlying my IV design might not hold. To ameliorate this concern, I follow the logic in the previous subsection and carefully choose a set of comparable control firms.

I obtain changes to S&P 500 index constituents from Siblis. 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 599 treated firms, 697 control firms in the index and 2,436 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.[14]

The top right panel of Figure 2 shows the level of passive ownership for S&P 500 additions and matched control firms around the month of index rebalancing. Again, within each

[14] Russell 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).

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.[15]

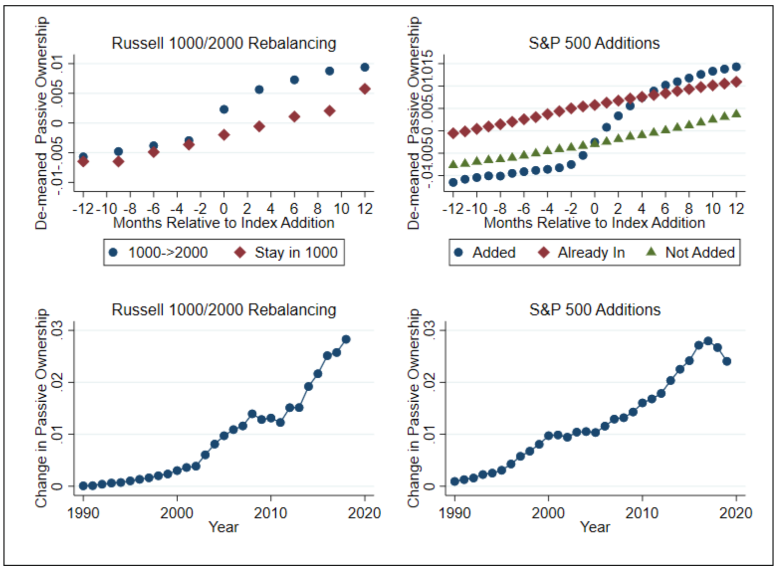

Figure 2. Effect of treatment on passive ownership. Top left panel: Average level of passive ownership for firms that stay in the Russell 1000 ("Stay in 1000") and firms that switched from the Russell 1000 to the Russell 2000 ("1000 $\rightarrow$ 2000"). Top right panel: Average level of passive ownership 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.

The two bottom panels of Figure 2 show the average change in passive ownership for

[15] S&P 500 index additions do not always coincide with the end of a calendar quarter. Given that the S12 data I use to quantify passive ownership is quarterly, I do not 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.

treated firms between month $t = -3$ and month $t = 3$ relative to the index reconstitution. The bottom left panel shows that for Russell 2000 switchers, the increase grew from almost nothing in 1990 to about 3% by 2019. Although average passive ownership has been growing over my sample, this is not sufficient for the change in passive ownership associated with switching from the Russell 1000 to the Russell 2000 to increase. Importantly, the passive ownership share of the Russell 2000 has been increasing at a *faster rate* than the passive ownership share of the Russell 1000.

The bottom right panel shows that the change in passive ownership from being added to the S&P 500 exhibits a similar upward trend. This increase can mostly be explained by the growing passive ownership share of the S&P 500. A notable deviation from this trend is the decline in the treatment effect for the last few years of the sample (i.e., the increase in passive ownership associated with being added to the S&P 500 shrunk from 2017 to 2019). This is due to two factors: (1) an increase in the share of S&P 500 additions which are migrations from the S&P MidCap 400 and (2) an increase in the passive ownership share of S&P MidCap-tracking funds (Greenwood and Sammon, 2022).

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 Passive $\text{Gap}_{i,t}$. For the Russell switchers, it is defined as the difference in passive ownership between firms in the Russell 1000 and the Russell 2000 within $\pm$ 100 ranks of the 1000th ranked firm in March (the last S12 filing date before index rebalancing). For the S&P 500 additions, Passive $\text{Gap}_{i,t}$ is the difference in passive ownership between the matched control firms in the index and out of the index, three months before the treated firm is added to the index. If at the time of index addition there are not matched control firms *both* in and out of the index, I use the average Passive $\text{Gap}_{i,t}$ 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 Passive $\text{Gap}_{i,t}$ to instrument for passive ownership. The two key pieces of the IV are therefore: (1) the instrumented change in passive ownership (2) the IV specification:

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

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

where $Outcome_{i,t}$ is $|Ret|$, $|Ret|/SD$, PJ or IVD and $Post_{i,t}$ 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)). Passive $\text{Gap}_{i,t} \times Treated_{i,t}$ 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 3 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. The coefficient on $Post \times Treated \times PassiveGap$ 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 $|Ret|$ on the left hand side. The effect of passive ownership on $|Ret|$ is positive, consistent with the cross-sectional regression results. Although the IV has a Local Average Treatment Effect (LATE) interpretation, it is still useful to compare the magnitude of the IV coefficients with those that come from the OLS regressions in Table 2. The IV estimate of 0.16 is about 2.5 times the OLS estimate of 0.0635. In column 3, the analogue to column 2 for $|Ret|/SD$, the IV estimate of 11.68 is also positive and about 2.5 times the OLS regression coefficient of 4.425. Column 4 shows that the IV estimate for PJ is positive and about 3.5 times the OLS estimate of 0.248. Finally, column 5 shows that the IV estimate for IVD is about 3 times as large as the OLS estimate in Table 2 of 0.112. 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 larger in magnitude than the OLS estimates suggest that the latter are not materially biased by this endogeneity concern.

I report the reduced form regressions i.e., regressions of the pre-earnings announcement price informativeness measures on the instruments themselves in Appendix Table D3. Although the reduced-form coefficient on $Post_{i,t}$ is almost always positive and statistically significant, the coefficient on Passive $\text{Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t}$ is not always significant. 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 Passive $\text{Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t}$ uses the expected *change* in passive ownership from index changes (i.e., Passive $\text{Gap}_{i,t}$), 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 Appendix D.4.

Panel B of Table 3 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. Something that stands out from the first stage regression is that the coefficient on the interaction term, Post $\times$ Treated $\times$ Passive Gap, is less than 1. One reason for this is that some firms added to the S&P 500 are migrations from the S&P MidCap. In these cases, buying by passive S&P 500 funds may be matched by selling from S&P MidCap funds, leading to an overall smaller than expected change in passive ownership, relative to e.g., a firm added to the S&P 500 from outside the S&P 1500 universe.

Columns 2 to 5 are the IV regressions, which all show a positive and statistically significant relationship between passive ownership and the share of information incorporated into prices after it is formally announced. Like Panel A, these point estimates are larger in magnitude than the cross-sectional regression estimates by a factor of about 3-4. 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.[16]

[16] Suppose 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.

	First Stage	Instrumental Variables			
	Panel A: Russell				
	Passive	$\lvert Ret \rvert$	$\lvert Ret \rvert / SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)
Post × Treated	1.77***				
× Passive Gap	(0.152)				
Post	0.02***				
	(0.001)				
Passive		0.16***	11.68***	0.90**	0.36***
		(0.037)	(2.352)	(0.417)	(0.088)
Observations	33,293	33,293	33,293	12,575	11,731
F-Statistic	222.7				

	First Stage	Instrumental Variables			
	Panel B: S&P 500				
	Passive	$\lvert Ret \rvert$	$\lvert Ret \rvert / SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)
Post × Treated	0.593***				
× Passive Gap	(0.060)				
Post	0.018***				
	(0.001)				
Passive		0.221***	14.744***	1.568***	0.413**
		(0.054)	(1.656)	(0.339)	(0.186)
Observations	262,893	262,893	262,893	98,111	142,249
F-Statistic	242				

Table 3 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness. Estimates from:

$$\begin{aligned} Passive_{i,t} = &\quad \alpha + \beta_1 Post_{i,t} + \beta_2 \text{Passive Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t} + FE + \epsilon_{i,t} \\ \text{Price informativeness}_{i,t} = &\quad \alpha + \beta_3 \widehat{Passive}_{i,t} + FE + \epsilon_{i,t} \end{aligned}$$

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$ 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 2-4 are instrumental variables 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.

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 changes 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.[17] To further alleviate the concern that institutional ownership is driving my results, in Appendix Table D4, I show the IV results are quantitatively unchanged by including the non-passive institutional ownership ratio on the right-hand side of both the first stage and the IV regressions.

Another threat to the exclusion restriction is that index switching is correlated with investor attention or learning, and these changes are what is actually driving the observed change in pre-earnings announcement price informativeness. In Appendix Table D7, I show that the two index inclusion experiments have different effects on analyst coverage. Specifically, stocks switching to the Russell 2000 receive less coverage, both in terms of the number of analysts and average analyst accuracy. On the other hand, stocks added to the S&P 500 on average have an increase in analyst coverage and accuracy. This is perhaps unsurprising, as firms switching to the Russell 2000 are shrinking, while firms being added to the S&P 500 are growing.

As another way to quantify investor learning/attention, Appendix Table D8 examines the effect of index switching on institutional investor attention (Ben-Rephael et al., 2017) and downloads of SEC filings (Loughran and McDonald, 2017). Empirically, being added to the S&P 500 is weakly associated with more downloads of SEC filing and Bloomberg terminal searches, while the effect of switching to the Russell 2000 on these measures of investor attention is mixed.

The results on analyst coverage and investor attention makes the consistency between the two experiments – in terms of the effect of passive ownership on pre-earnings announcement price informativeness – more surprising. For stocks added to the S&P 500, investors

[17] A 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.

and analysts are paying more attention, while for stocks switching from the Russell 1000 to the Russell 2000, there is evidence of the opposite. In both cases, however, the instrumented change in passive ownership predicts decreases in pre-earnings announcement price informativeness. This suggests that differences in investor attention (i.e., a violation of the only-through assumption) are likely not driving my IV results.

Yet another threat to the exclusion restriction is that index switching changes the nature of the earnings announcements themselves. One specific possibility is that firms change the way they disclose information to investors after they are e.g., added to the S&P 500. To allay this concern, in Appendix Table D8, I show that being added to the S&P 500 or switching to the Russell 2000 has almost no effect on measures of 10-K complexity (Loughran and McDonald, 2020).

Similarly, it's possible that being added to an index changes the volatility of earnings news, and therefore the difficulty of the associated learning problem. For example, if earnings became more volatile after a stock switched from the Russell 1000 to the Russell 2000 (e.g., because the firm is shrinking), one might expect price informativeness to decline. Appendix Table D7 shows that switching to the Russell 2000 is not associated with more volatile earnings growth, suggesting this is not a key driver of the IV results for the Russell experiment. There is an increase in earnings volatility for firms added to the S&P 500, but as I discuss in Appendix D.7, this could be due the earnings-gaming some firms engage in to be added to the index.[18]

Outside of possible violations of the only-through assumption, a final concern with the results in Table 3 is that many previous studies have used switching between 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

[18] An example of this is Tesla, which was large enough to be added to the S&P 500, but did not meet S&P's profitability thresholds. One contribution to Tesla becoming profitable enough to be added to the S&P 500 was the sale of regulatory credits, which the company knew was not going to be a sustainable source of profit.

for how many times the natural experiment has been re-used. Table 3 shows that almost all 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 IV based on Russell switchers yields similar point estimates to the the IV based on S&P addition, 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.

5 Conclusion

In this paper, I propose several 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 regressions implies that a 15% increase in passive ownership decreases pre-earnings announcement price informativeness by $1/4^{th}$ of its whole sample mean and $1/6^{th}$ of its whole sample standard deviation. To establish causality, I show that these results continue to hold when restricting to quasi-exogenous variation in passive ownership associated with switching into the Russell 2000 and being added to the S&P 500.

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 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

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

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).

Baker, S. R., Bloom, N., Davis, S. J., Kost, K., Sammon, M., and Viratyosin, T. (2020). The unprecedented stock market reaction to covid-19. *The review of asset pricing studies*, 10(4):742–758.

Ball, R. and Brown, P. (1968). An Empirical Evaluation of Accounting Income Numbers. *Journal of Accounting Research*, 6(2).

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.

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.

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

Chung, K. H. and Zhang, H. (2011). Corporate governance and institutional ownership. *Journal of financial and quantitative analysis*, 46(1):247–273.

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.

Crane, A. D. and Crotty, K. (2018). Passive versus active fund performance: do index funds have skill? *Journal of Financial and Quantitative Analysis*, 53(1):33–64.

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

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.

Duarte, J., Hu, E., and Young, L. (2020). A comparison of some structural models of private information arrival. *Journal of Financial Economics*, 135(3):795–815.

Dubinsky, A., Johannes, M., Carr, P., Chen, J., Chernov, M., Duffie, D., Glasserman, P., Heidari, M., Hull, B., Jagannatahn, R., Jones, C., Knight, J., and Linnainmaa, J. (2006). Fundamental uncertainty , earning announcements and equity options. *Wall Street Journal*, 1(August 2015).

Ernst, T. (2020). Stock-Specific Price Discovery From ETFs. *Working Paper*, 1(October 2017).

Frazzini, A. (2006). The disposition effect and underreaction to news. *The Journal of Finance*, 61(4):2017–2046.

Ganuza, 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).

Greenwood, R. and Sammon, M. C. (2022). The disappearing index effect. Technical report, National Bureau of Economic Research.

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).

Hartzmark, S. M. and Shue, K. (2018). A tough act to follow: Contrast effects in financial markets. *The Journal of Finance*, 73(4):1567–1613.

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

Kacperczyk, M. T., Nosal, J. B., and Sundaresan, S. (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., Li, J., Li, X., and Sheng, M. (2023). Etf activities and analysts' forecasts. *Available at SSRN 4398547*.

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.

Loughran, T. and McDonald, B. (2020). Measuring firm complexity. *Journal of Financial and Quantitative Analysis*, pages 1–55.

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*.

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.

Velury, U. and Jenkins, D. S. (2006). Institutional ownership and the quality of earnings. *Journal of business research*, 59(9):1043–1051.

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.

Zou, Y. (2018). Lost in the rising tide: Etf flows and valuation.

Internet Appendix for Passive Ownership and Price Informativeness

A Data details

A.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.
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.

A.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 Appendix C.2.1 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).

A.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. I use the WRDS MF LINKS database to connect the funds identified as passive in CRSP (using the method in Appel et al. (2016)) 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. My overall numbers for passive ownership closely mirror the Investment Company Institute (ICI) factbook (Figure 6.6).

Finally, 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 my results are quantitatively unchanged if I instead fix passive ownership at its last reported level between the ends of calendar quarters.

A.4 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 least 5 days after τ.[19]

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

[19]This 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.

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:

$$\overline{IV}_{i,b} = Mean\left[IV_{i,(b-s,b),b} : b-s \in [\tau-20, \tau-1]\right] \tag{A1}$$

$\overline{IV}_{i,a}$ 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:

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right) \tag{A2}$$

higher values of $IVD_{i,\tau}$ imply that options which span earnings announcements are relatively more expensive.[20] 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.[21]

Figure A.1 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 increase in $|Ret|$), on both an equal-weighted and value-weighted basis, IVD has increased by about 5 over the past 25 years.

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

[21] Suppose 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 $\overline{IV}_{i,c}$.

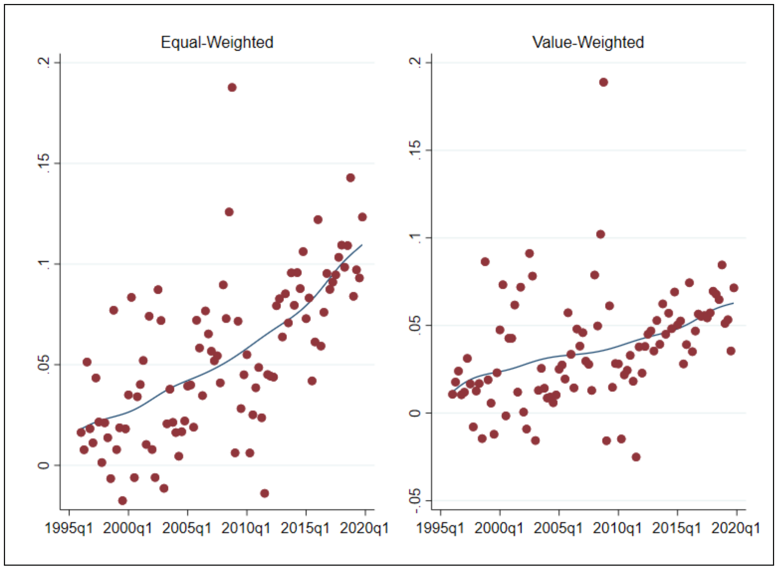

Figure A.1. Time-series trends in IVD. Equal-weighted and value-weighted averages of IVD by quarter. Red dots represent cross-sectional averages and blue lines represent LOWESS filters with bandwidths equal to 20% of quarters in the dataset.

B Stylized Facts

B.1 Distributions of and relationship between measures of pre-earnings announcement price informativeness

In Figure B.2, I plot histograms of each of the four measures of pre-earnings announcement price informativeness. These have all been Winsorized at the 1% and 99% level, which is why there appears to be a large mass at either end of the histograms. The top two panels show that $|Ret|$ and $|Ret|/SD$ seem to have exponential distributions. The bottom two panels show that PJ and IVD have something closer to normal distributions, except both have a slight positive skew.

Next, I turn to the correlation among the measures of price informativeness. Figure B.3 presents bin scatter plots of each measure of pre-earnings announcement price informativeness against each other measure. Generally speaking, there is a positive relationship between all the measures of pre-earnings announcement price informativeness. The strongest relationship is between $|Ret|$ and $|Ret|/SD$, while the weakest is between IVD and all the other measures. This is likely because $|Ret|$, $|Ret|/SD$ and PJ are functions of what actually

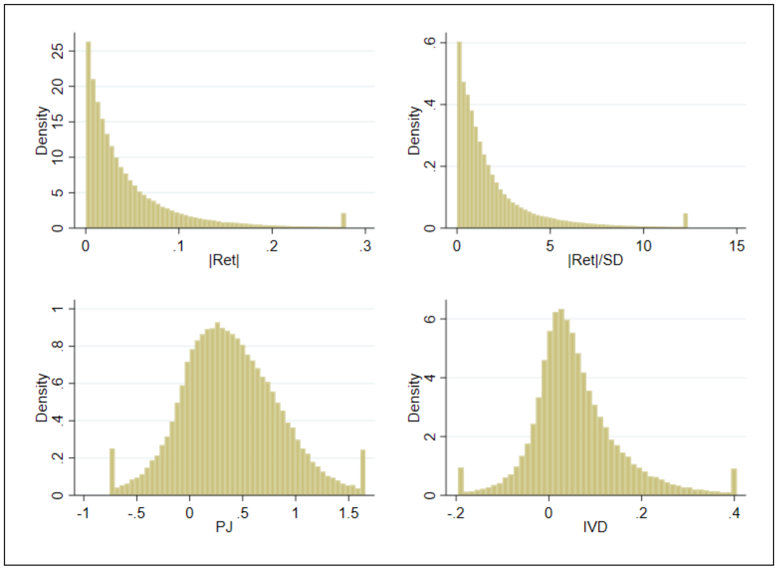

Figure B.2. Distributions of $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ and $IVD_{i,t}$. Histograms of $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ and $IVD_{i,t}$. All variables are Winsorized at the 1% and 99% level. Observations are equal weighted.

occurred on the earnings announcement day, while the construction of IVD only uses option prices from up until the day before the earnings announcement.

A striking feature of this plot is the difference between the top right panel ($|Ret|$ vs. IVD) and the bottom middle panel ($|Ret|/SD$ vs. IVD). The relationship between $|Ret|$ and IVD appears to be U-shaped, while it becomes S-shaped when comparing $|Ret|/SD$ to IVD. This is due to firms which have high $|Ret|$ having high volatility on average before/after the announcement as well, which will go both into realized volatility (i.e., SD the standard deviation of pre-earnings announcement returns) and implied volatility for options with surrounding expiration dates.

B.2 Relationship between passive ownership and firm-level characteristics

Figure B.4 presents bin-scatter plots with the 9 control variables from the baseline regressions (Equation 4) on the y-axis and passive ownership on the x-axis. To make quantities comparable across time, all variables have been normalized to have mean zero and standard

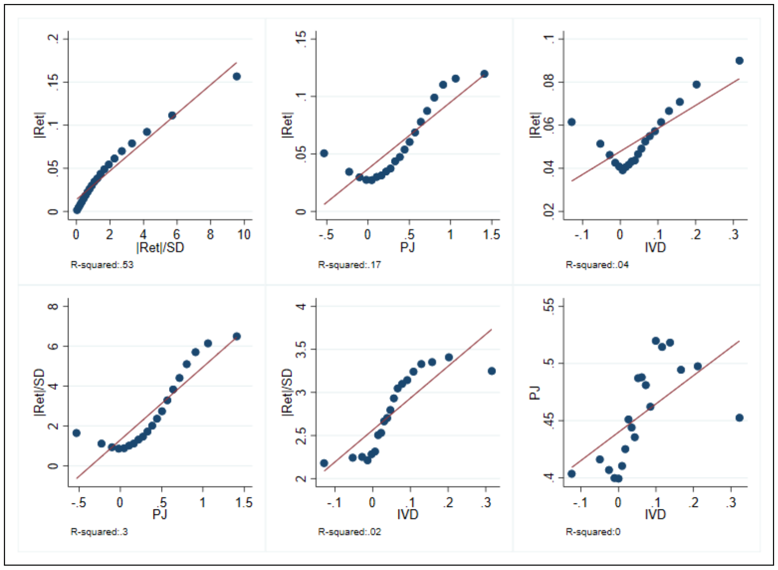

Figure B.3. Relationship between $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ and $IVD_{i,t}$. Bin-scatter plots (binned into vingtiles) and best fit lines between $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ and $IVD_{i,t}$. R-squared is from a univariate regression of the variable on the y-axis on the variable on the x-axis. Observations are equal weighted.

deviation one each quarter. Some of the variables have a strong relationship with passive ownership, like total institutional ownership (which is unsurprising, as all mutual funds are included in institutional ownership) and CAPM R-squared (although that may be an effect of passive ownership itself, see e.g., Israeli et al. (2017) and Bennett et al. (2020b)). Other variables are almost totally unrelated to passive ownership. For example, and perhaps surprisingly, the average relationship between passive ownership and market capitalization is not very strong (R-squared of only 0.02). That being said, the strength of this relationship varies significantly over time, as Figure D.7 shows.

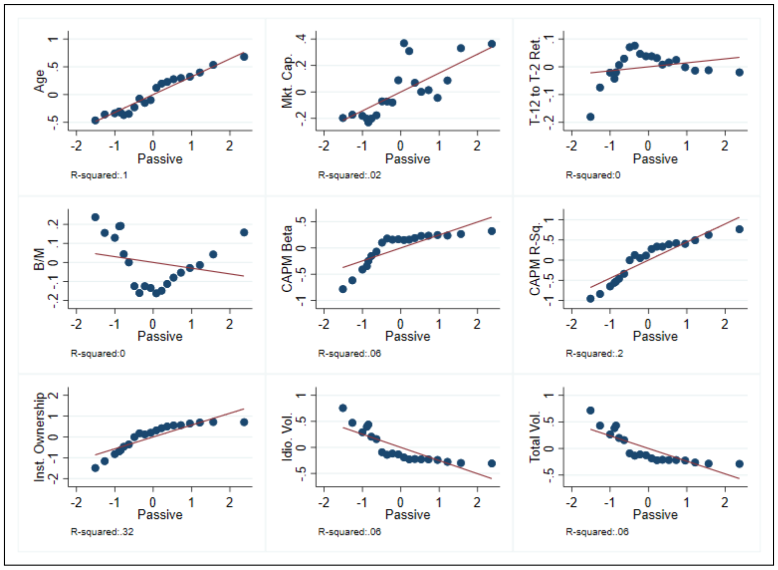

Figure B.4. Relationship between passive ownership and firm characteristics. Binscatter plots (binned into vingtiles) and best fit lines between the firm-level control variables and passive ownership. All variables are normalized to have a mean of zero and a standard deviation of one each quarter. Observations are equal weighted. R-squared is from a univariate regression of the variable on the y-axis on passive ownership.

C Cross-Sectional Regressions

C.1 Time variation in relationship between passive ownership and pre-earnings announcement price informativeness

To better understand time series variation in the effect of passive ownership on pre-earnings announcement price informativeness, I run the following regression:

$$\text{Prc. Inf.}_{i,t} = \sum_{\tau=1990}^{2018} \beta_\tau 1_{year=\tau} + \sum_{\tau=1990}^{2019} \alpha_\tau 1_{year=\tau} \times Passive_{i,t} + \gamma X_{i,t} + \phi_q + \psi_i + e_{i,t} \quad \text{(C3)}$$

where ϕ_q are quarter-of-year fixed effects, designed to account for seasonality. Note that 2019 is the omitted year fixed effect. I run this pooled regression, rather than running the regression year-by-year, so the results are most analogous to the regressions in Table 2, because the estimated firm fixed effects and coefficients on the control variables will be nearly identical.

I plot the coefficients on the interaction terms, α_τ in Figure C.5. The horizontal red line represents the point estimate a regression of pre-earnings announcement price informativeness on passive ownership, pooling across all years. The first striking fact in this Figure is that in the early years, there is significant variation in the estimated effects. One reason for this is that passive ownership was tiny (e.g., SPY, a very popular S&P 500 ETF wasn't launched until 1993, and IWB and IWM, the largest Russell 1000 and 2000 ETFs weren't launched until 2000), which may be driving the extreme variation in the estimated coefficients. I would also like to highlight that because the regressions in Figure C.5 are equal weighted and there are fewer observations in the early years, the data in those years may not contribute much to the overall pooled OLS estimate. After these early years, however, the effect seems to stabilize around the pooled regression estimate.

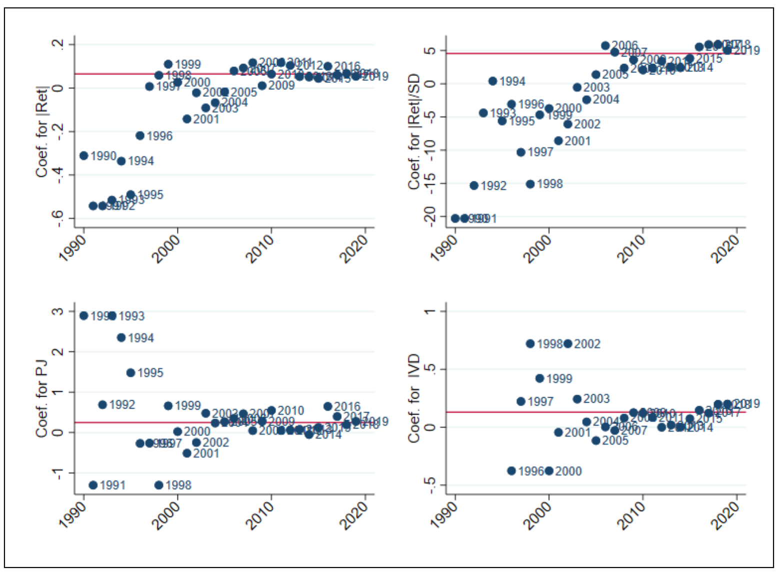

Figure C.5. Cross-sectional regression of price informativeness on passive ownership by year. Figure with estimates of α_τ from:

$$\text{Prc. Inf.}_{i,t} = \sum_{\tau=1990}^{2018} \beta_\tau 1_{year=\tau} + \sum_{\tau=1990}^{2019} \alpha_\tau 1_{year=\tau} \times Passive_{i,t} + \gamma X_{i,t} + \phi_q + \psi_i + e_{i,t}$$

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$. 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.

C.2 Alternative explanations for the negative relationship between passive ownership and pre-earnings announcement price informativeness

In this subsection, I discuss two threats to identification in my baseline regressions (1) Regulation Fair Disclosure and (2) the rise of algorithmic trading.

C.2.1 Regulation Fair Disclosure (Reg FD)

Before Reg FD was passed in August, 2000, firms would disclose earnings information to selected analysts before it became public. This information 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 all the measures of pre-earnings announcement price informativeness, as there was a large negative shock to the amount of information firms released before earnings announcements after it was passed. All the measures, 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 Table C1, I re-run the cross-sectional regressions using only post-2000 data. The point estimates are quantitatively similar, which alleviates concerns of my results being driven by Reg FD.

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

Weller (2018) shows that Algorithmic Trading (AT) activity is negatively correlated with pre-earnings announcement 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 2012-2019. 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 2019. 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. Following 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 in Table 2 is almost entirely the result of restricting to data between 2012 and 2019.

Table C1 replicates the baseline OLS regressions on the matched sample to MIDAS data in columns 1, 3, 5 and 7. Columns 2, 3, 6 and 8 add 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.

	Controlling for Algorithmic Trading Activity								Post Reg-FD			
	$\|Ret\|$	$\|Ret\|$	$\|Ret\|/SD$	$\|Ret\|/SD$	PJ	PJ	IVD	IVD	$\|Ret\|$	$\|Ret\|/SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Passive	0.0146	0.0199	3.264***	2.951***	0.236**	0.203*	0.117***	0.110***	0.0463***	3.913***	0.215***	0.100***
	(0.0143)	(0.0139)	(0.5950)	(0.6080)	(0.1030)	(0.1070)	(0.0399)	(0.0398)	(0.0080)	(0.4000)	(0.0624)	(0.0317)
Cancel to Trade		-0.00477***		-0.157***		-0.0362***		-0.0118**				
		(0.0011)		(0.0393)		(0.0123)		(0.0048)				
Trade to Order		-0.00113		-0.269***		-0.144***		-0.00521				
		(0.0012)		(0.0579)		(0.0141)		(0.0051)				
Odd Lot Ratio		-0.00461***		-0.0304		0.0312**		-0.00764*				
		(0.0012)		(0.0481)		(0.0136)		(0.0043)				
Avg. Trade Size		0.00138		-0.573***		-0.0504***		0.0125				
		(0.0019)		(0.0724)		(0.0179)		(0.0085)				
Observations	91,532	91,532	91,532	91,532	37,778	37,778	41,393	41,393	267,035	267,035	99,512	89,629
R-squared	0.23	0.232	0.208	0.21	0.19	0.197	0.352	0.353	0.19	0.191	0.158	0.308

Table C1 Corroborating evidence for effect of passive ownership on pre-earnings announcement price informativeness.
Table with estimates of β from:

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

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$. 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. Columns 1-8 restrict to observations that can be matched to the SEC MIDAS data. Columns 2, 4, 6 and 8 include controls for the AT measures in Weller (2018). Standard errors double clustered at the firm and year-quarter level in parenthesis. Columns 9-12 restrict to observations between 2001 and 2019.

C.3 Robustness to alternative definitions of passive ownership

As discussed above, the CRSP index fund flag is not well populated early in my sample, which is one reason I favor a name-based classification system over one based purely on using CRSP's index fund flag. A possible downside to this, however, is misclassification (Crane and Crotty, 2018). That being said, I find that the level of passive ownership is nearly identical using their definition (strictly focusing on funds with an index fund flag of D in the CRSP mutual fund database) as the name-based definition of Appel et al. (2016). Further, both methods yield numbers close to those published by the Investment Company Institute. For example, in 12/2022, index fund ownership is 17.1% of the market under the Appel et al. (2016) definition, while is it 16.7% under the Crane and Crotty (2018) definition.

This similarity is not specific to data from recent years. Figure C.6 shows that there is almost no difference between the Appel et al. (2016) definition of passive ownership (blue line) and the stricter definition in Crane and Crotty (2018) (red line).

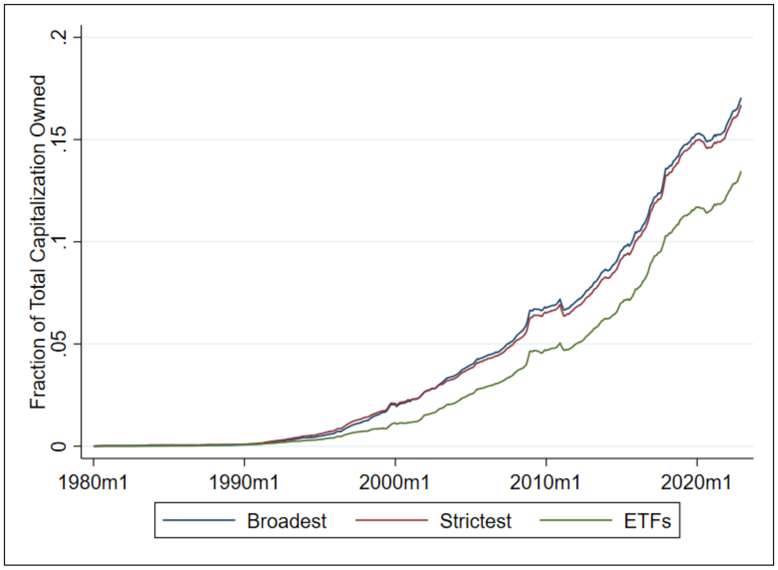

Figure C.6. Trends in Passive Ownership. Fraction of total capitalization of ordinary common shares traded on major exchanges owned by passive funds and ETFs. Broadest includes funds (A) classified as any type of index fund by the CRSP mutual fund database (i.e., an index-based fund, a pure index fund or an enhanced index fund) or (B) which have a name that makes them look like an index fund (following Appel et al. (2016)). Strictest includes only funds with an index fund flag of "D" in the CRSP mutual fund database. ETFs includes only funds with an ETF flag of "F" in the CRSP mutual fund database.

I also find that using this alternative definition of passive ownership (i.e., only including those funds with an index fund flag of "D") has essentially no effect on my main results. In Table C2, columns 1-4 replicate the OLS regression results in Table 2. Columns 5-8 mirror columns 1-4, except they use the strict definition of passive ownership from Crane and Crotty (2018). If anything, the effects are slightly larger using their stricter definition. This, however, could be mechanical, in that it excludes some types of passive ownership which are correlated with strict index fund ownership, but which also decrease pre-earnings announcement price informativeness.

	$\|Ret\|$ (1)	$\|Ret\|/SD$ (2)	PJ (3)	IVD (4)	$\|Ret\|$ (5)	$\|Ret\|/SD$ (6)	PJ (7)	IVD (8)	$\|Ret\|$ (9)	$\|Ret\|/SD$ (10)	PJ (11)	IVD (12)
Broad Passive	0.0635***	4.425***	0.248***	0.112***								
	(0.007)	(0.384)	(0.060)	(0.030)								
Strict Passive					0.0700***	5.610***	0.252***	0.143***				
					(0.008)	(0.401)	(0.066)	(0.031)				
ETFs									0.0745***	5.842***	0.271***	0.160***
									(0.009)	(0.459)	(0.077)	(0.036)
Observations	441,238	441,238	148,864	111,604	441,238	441,238	148,864	111,604	441,238	441,238	148,864	111,604
R-squared	0.194	0.214	0.163	0.3	0.195	0.215	0.163	0.3	0.194	0.215	0.163	0.3
Weight	Equal	Equal	Equal	Equal	Value	Value	Equal	Equal	Equal	Equal	Value	Value
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Fixed Effects	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ

Table C2 Cross-sectional regression of price informativeness on different types of passive ownership. Table with estimates of β from:

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

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$. 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. Broadest definition of passive includes funds (A) classified as any type of index fund by the CRSP mutual fund database (i.e., an index-based fund, a pure index fund or an enhanced index fund) or (B) which have a name that makes them look like an index fund (following Appel et al. (2016)). Strictest definition of passive includes only funds with an index fund flag of "D" in the CRSP mutual fund database. ETFs includes only funds with an ETF flag of "F" in the CRSP mutual fund database. Standard errors double clustered at the firm and year-quarter level in parenthesis.

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 D.7 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% (the relationship between passive ownership and market capitalization has become stronger over time, which is why this is higher than the R-squared reported in B.4). 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 a relatively larger passive ownership share than large-cap indices (e.g. the Russell 1000) (Pavlova and Sikorskaya, 2022).

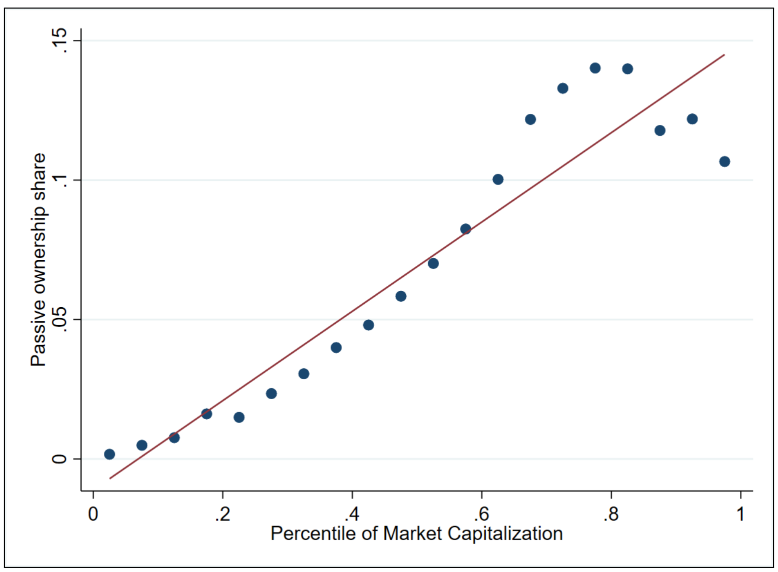

Figure D.7. Passive ownership and percentile of market capitalization. Data from 12/2018. Includes all firms with both non-missing passive ownership and non-missing market capitalization.

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 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.[22]
- 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 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.

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

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 ± 2.5% of the total market capitalization of the Russell 3000.[23]
- 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 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

[23]In reality, the bands are ± 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.

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 D.8 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.

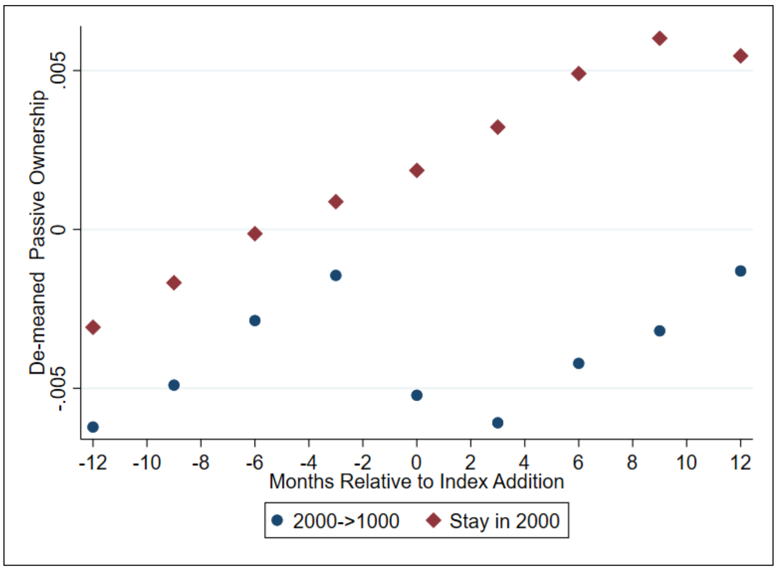

Figure D.8. Russell 1000/2000 Reconstitution and Changes in Passive Ownership. Average level of passive ownership for firms that stay in the Russell 2000 (control firms) and firms that moved from the Russell 2000 to the Russell 1000 (treated firms). Passive ownership is demeaned within each cohort.

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 Reduced form regressions

I report the reduced form regressions in Table D3. One concern with these results is that the IV results in Table 3 are always significant, while the reduced form coefficient on the interaction term between Post, Treated and Passive Gap is almost always insignificant. The worry 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 both experiments). In the next subsection, 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.[24] I also discuss how the presence of shadow indexing (Mauboussin, 2019) could explain why the IV is significant but the reduced form is not.

D.5 Statistical significance of instrumental variables vs. reduced form

In Table 3, the IV regressions are highly significant, while in Table D3 the reduced form regressions show a consistently insignificant coefficient on one of the two instruments. 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:

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

[24]The 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 3), it is possible for the IV to be significant but the reduced form to be insignificant.

	Panel A: Russell			
	$\lvert Ret \rvert$	$\lvert Ret \rvert/SD$	PJ	IVD
	(1)	(2)	(3)	(4)
Post × Treated	0.18	1.61	-0.31	0.93***
× Passive Gap	(0.121)	(6.955)	(1.548)	(0.297)
Post	0.00***	0.31***	0.03***	0.01*
	(0.001)	(0.049)	(0.010)	(0.003)
Observations	33293	33293	12575	11731
R-squared	0.268	0.253	0.207	0.304

	Panel B: S&P			
	$\lvert Ret \rvert$	$\lvert Ret \rvert/SD$	PJ	IVD
	(1)	(2)	(3)	(4)
Post × Treated	0.053	1.513	0.779	-0.138
× Passive Gap	(0.063)	(4.128)	(0.711)	(0.119)
Post	0.004***	0.278***	0.032***	0.010**
	(0.001)	(0.028)	(0.007)	(0.004)
Observations	263348	263348	98242	142318
R-squared	0.285	0.281	0.237	0.308

Table D3 Reduced form estimates for effect of passive ownership on pre-earnings announcement price informativeness. Estimates from:

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

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$ 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. 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.

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.[25] At a high level, this is likely not a problem in my setting, as the first stage is very strong ($F > 200$ for both experiments).

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

$$\begin{aligned} \text{Structural} \quad &: \quad y_i = \beta x_i + \varepsilon_i && \text{(D4)} \\ \text{First stage} \quad &: \quad x_i = \gamma z_i + u_i && \text{(D5)} \\ \text{Reduced Form} \quad &: \quad y_i = \overbrace{\beta\gamma}^{\alpha} z_i + \overbrace{\beta u_i + \varepsilon_i}^{v_i} && \text{(D6)} \end{aligned}$$

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:

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

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

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

Writing out the definition of the IV estimator:

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

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

Unlike the OLS estimator, $\widehat{\beta}_{IV}$ 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:

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

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

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

Like the IV estimator, $\widehat{\alpha}_{RF}$ 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:

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

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

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

And

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

Thus, their joint distribution is:

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

where

$$\rho_{\varepsilon,v} = \frac{E\left[v_i \varepsilon_i\right]}{\sqrt{\left(E\left[v_i^2\right]\right)\left(E\left[\varepsilon_i^2\right]\right)}} = \frac{\beta \frac{E[u_i \varepsilon_i]}{E\left[\varepsilon_i^2\right]} + 1}{\sqrt{\beta^2 \frac{E\left[u_i^2\right]}{E\left[\varepsilon_i^2\right]} + 2\beta \frac{E[u_i \varepsilon_i]}{E\left[\varepsilon_i^2\right]} + 1}} \tag{D7}$$

Equation D7 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:

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

and taking their ratio yields:

$$\frac{t_{\widehat{\beta}_{IV}}}{t_{\widehat{\alpha}_{RF}}} = \frac{\sqrt{\left(N^{-1} \sum_i \widehat{v}_i^2\right)}}{\sqrt{\left(N^{-1} \sum_i \widehat{\varepsilon}_i^2\right)}} \xrightarrow{p} \sqrt{\frac{E\left[v_i^2\right]}{E\left[\varepsilon_i^2\right]}} = \sqrt{\frac{\beta^2 E\left[u_i^2\right] + 2\beta E\left[u_i \varepsilon_i\right] + E\left[\varepsilon_i^2\right]}{E\left[\varepsilon_i^2\right]}} \tag{D8}$$

Thus, under the $\beta = \alpha = 0$ null, these t-statistics will be perfectly correlated asymptotically.[26]

Empirically, in Table 3, I am testing whether $\widehat{\beta}_{IV} = 0$. In Table D3 I am testing whether $\widehat{\alpha}_{RF} = 0$, although, as I discuss below, because I have multiple instruments, the interpretation is slightly different.

With this in mind, assuming the true β and α are not zero, there are three things to consider:

[26]If $\beta = \alpha = 0$ 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.

1. Equation D8 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, $t_{\widehat{\beta}_{IV}} = 2.5$ so the IV is statistically significant, the true β is 0.5, and $E\left[u_i^2\right] = E\left[\varepsilon_i^2\right] = 1$. Then, if the covariance between u_i and ε_i is higher than 0.4, the reduced form coefficient will not be both positive and significant at the 5% level. This is because with these parameters $\sqrt{\frac{\beta^2 E\left[u_i^2\right]+2\beta E[u_i\varepsilon_i]+E\left[\varepsilon_i^2\right]}{E\left[\varepsilon_i^2\right]}} = 1.28$ and $2.5/1.28 < 1.96$.
2. More generally, in my setting I expect $\beta > 0$ i.e., passive ownership decreases pre-earnings announcement price informativeness (recall that $\beta > 0$ implies a larger share of the earnings information is incorporated into prices *after* the announcement itself). If this is the case, as the covariance between u_i and ε_i becomes more positive, 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 covariance will tend to increase $\beta E\left[u_i\varepsilon_i\right]$ in the numerator of Equation D8, 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, under the assumption that the true $\beta > 0$, the structural regression will undershoot the true change in price informativeness (i.e., the structural equation residual ε_i will be positive), leading to a positive 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 $\psi > 0$.

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 (ignoring the fixed effects for now):

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

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 positive covariance (recall that larger values of β imply less informative pre-earnings announcement prices), 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 D8 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 D4 (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 $\hat{\sigma}^2_\varepsilon$ and $\hat{\sigma}^2_v$ depend on N), I choose $N = 30,000$ to match the number of observations in Panel A of Table 3. 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 D.9 plots the fraction of simulations where the t-statistic from the IV is greater than 1.96, but the t-statistic from the RF is less 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 D8 will be bigger than one.

The red dots show that as the correlation between u_i and ε becomes more negative, the RF becomes more statistically significant on average. This is because the numerator in Equation D8 shrinks, as this negative covariance between u_i and ε is being multiplied by β, which is greater than 0. The blue dots show that as the correlation between u_i and ε becomes more positive, the RF becomes less significant on average than the IV. In this case, the positive covariance between u_i and ε is being multiplied by the positive beta, which increases the numerator of Equation D8.

Multiple Instruments

Before moving on, I want to highlight that in my setting, the exact relationship between the IV and the reduced form is more complicated than the algebra above would imply. The issue is that in my setting, I have two instruments for passive ownership, while in Lochner and Moretti (2004), they only have a single instrument. Recall that in Table 3, I am using both $Post_{i,t}$ and Passive $\text{Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t}$ as instruments for passive ownership. Over my sample, there is a time trend toward increased passive ownership.

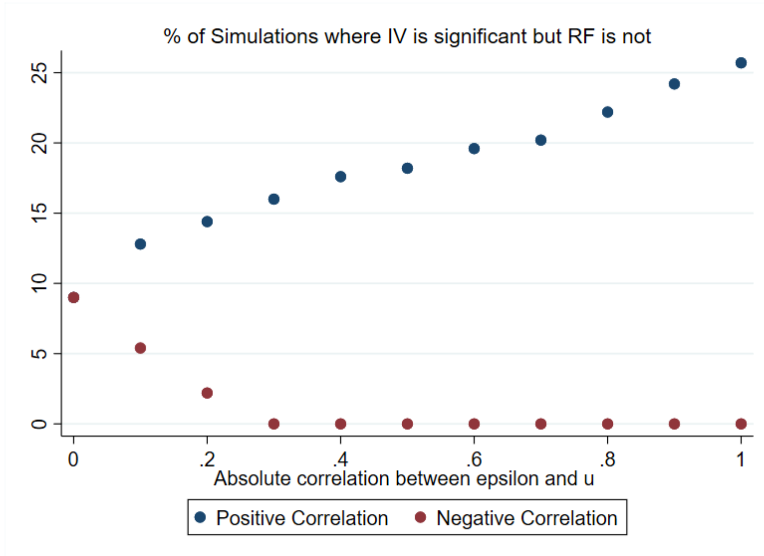

Figure D.9. Comparison of statistical significance. Each dot represents the percentage of simulations where the instrumental variables specification is statistically significant, but the reduced form is not. The blue dots are from simulations where ε positively correlated with u, while the red dots are from simulations where ε is negatively correlated with u. Moving from left to right increases the (absolute) correlation between ε and u.

This is why in the first stage regression in Table 3, the coefficient on both $Post_{i,t}$ and Passive Gap$_{i,t} \times Treated_{i,t} \times Post_{i,t}$ are positive and statistically significant.

With that in mind, if passive ownership causes price informativeness to decline, we expect two things to be true in the reduced form regressions: (1) the coefficient on $Post_{i,t}$ to be positive and statistically significant (2) the coefficient on Passive Gap$_{i,t} \times Treated_{i,t} \times Post_{i,t}$ to be positive and statistically significant. Across 7 of the 8 reduced-form regressions, the coefficient on $Post_{i,t}$ is indeed positive and statistically significant. And, in the only regression where it is not, the coefficient on Passive Gap$_{i,t} \times Treated_{i,t} \times Post_{i,t}$ is positive and statistically significant, consistent with the 2nd prediction.

In 6 of the 8 regressions, the coefficient on Passive Gap$_{i,t} \times Treated_{i,t} \times Post_{i,t}$ is positive, as expected, but in 5 of these cases it is not statistically significant. In terms of magnitudes, in Panel A of Table D3, column 2 implies that a 4% expected increase in passive ownership after treatment (the expected increase in passive ownership for stocks switching from the Russell 1000 to the Russell 2000 in June 2019) implies an increase in $|Ret|/SD$ of 0.0644. This is above and beyond the average increase of 0.31 in $|Ret|/SD$ for all observations (i.e., both the treated and control firms) in the post period.

The coefficient on Passive $\text{Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t}$ is negative for PJ in the Russell experiment and IVD for the S&P experiment. While this is inconsistent with passive ownership decreasing pre-earnings announcement price informativeness, I would like to highlight that not only are these estimates statistically insignificant, but the magnitudes are small. For example, for PJ in the Russell experiment, a 4% expected increase in passive ownership implies a -0.012 decrease in PJ. This is relative to an average increase in the post period of 0.03.

D.6 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 non-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 is little change in overall institutional ownership (Gloßner (2018), Appel et al. (2020)).

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 D.10 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.

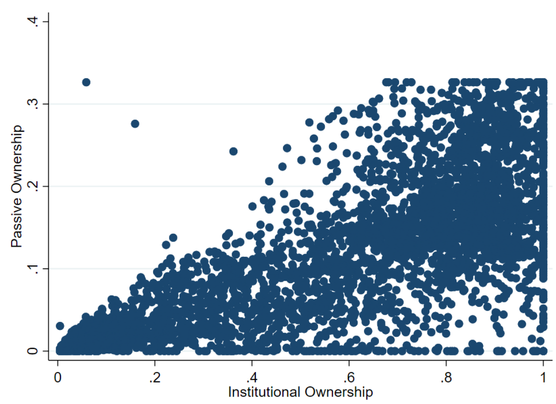

Figure D.10. Passive Ownership vs. Institutional Ownership. Plot of passive ownership against total institutional ownership in 12/2018. Both quantities are Winsorized at the 1% and 99% level.

The second piece of evidence is in Table D4, where I replicate all the instrumental variables regressions, including total non-passive institutional ownership on the right hand side. All the results are quantitatively unchanged from Table 3 in the main body of the paper, allaying concerns that institutional ownership is driving my results.

D.7 Other threats to the only-through assumption

Contemporaneous changes in institutional ownership are not the only threat to the only-through assumption which underlies my IV analysis. First, I examine the cross-sectional relationship between passive ownership on several quantities one might naturally expect to be correlated with pre-earnings announcement price informativeness: (1) analyst coverage (2) investor attention (3) fundamental earnings volatility and (4) complexity of earnings news.

In Table D5, I examine the cross-sectional relationship between passive ownership and various facets of analyst coverage. Column 1 shows that analyst coverage is negatively related to passive ownership. In terms of magnitudes, a 10% increase in passive ownership is correlated with roughly one fewer estimate on average, relative to a mean of 9 analysts and a standard deviation of 6 analysts. This mirrors results on the relationship between passive ownership and analyst coverage in Israeli et al. (2017) and Coles et al. (2022).

	Panel A: Russell				
		Instrumental Variables			
	Non-Pass Inst. Own	$\lvert Ret\rvert$	$\lvert Ret\rvert/SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)
Post × Treated	-2.58***				
× Passive Gap	(0.525)				
Post	0.04***				
	(0.005)				
Passive		0.15***	10.72***	0.94**	0.34***
		(0.037)	(2.289)	(0.428)	(0.089)
Observations	262,893	262,893	262,893	98,111	142,249
F-Statistic	43.76				

	Panel B: S&P 500				
		Instrumental Variables			
	Non-Pass Inst. Own	$\lvert Ret\rvert$	$\lvert Ret\rvert/SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)
Post × Treated	-0.206				
× Passive Gap	(0.405)				
Post	0.034***				
	(0.003)				
Passive		0.225***	14.348***	1.427***	0.412**
		(0.056)	(1.669)	(0.338)	(0.189)
Observations	262,893	262,893	262,893	98,111	142,249
F-Statistic	48.42				

Table D4 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness controlling for non-passive institutional ownership.
Estimates from:

$$\begin{aligned} Passive_{i,t} &= \alpha + \beta_1 Post_{i,t} + \beta_2 \text{Passive Gap}_{i,t} \times Treated_{i,t} \times Post_{i,t} + NPI_{i,t} + FE + \epsilon_{i,t} \\ \text{Price informativeness}_{i,t} &= \alpha + \beta_3 \widehat{Passive}_{i,t} + NPI_{i,t} + FE + \epsilon_{i,t} \end{aligned}$$

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$ 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. $NPI_{i,t}$ is non-passive institutional ownership, defined as all 13F ownership minus passive ownership. Column 1 in each panel is a regression of non-passive institutional ownership ($NPI_{i,t}$) on the instruments. Columns 2-4 are instrumental variables 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.

Column 2 shows that there is essentially no relationship between passive ownership and the standard deviation of analyst estimates. Column 3 examines the relationship between the definition of standardized unexpected earnings (SUE) from Hartzmark and Shue (2018), which is the difference between actual earnings and the mean estimate of earnings, divided by the pre-earnings announcement price. I take the absolute value of this, $|Dist.|$, before normalizing by the price because I am interested in whether the overall surprise is larger i.e., whether analysts covering the stock became less accurate. Here, the relationship is positive, consistent with high passive ownership stocks having larger earnings surprises. In terms of magnitudes, a 10% increase in passive ownership implies a 0.00087 larger SUE, relative to a mean of 0.0045 and a standard deviation of 0.0105.

Column 4 uses an alternative definition of SUE: the absolute difference between actual earnings and the mean estimate of earnings, divided by the standard deviation of analyst estimates. The logic behind using this alternative measure of SUE (as opposed to the measure in Hartzmark and Shue (2018)) is that passive ownership is correlated with pre-earnings announcement price informativeness. Dividing by the price (as in column 3), therefore, could conflate the effects of decreased analyst accuracy with decreased price informativeness. In column 4 the coefficient is again positive and statistically insignificant, suggesting decreased analyst accuracy for stocks with more passive ownership.

Next, I examine the standard deviation of year-over-year earnings growth. Specifically, for each firm, I compute year-over-year earnings growth using IBES earnings and then, following Novy-Marx (2015), compute the standard deviation of this earnings growth in 8 quarter rolling windows (results are similar using EPSFXQ from Compustat). Higher values, therefore, denote more volatile earnings growth. If earnings volatility were correlated with passive ownership, one might be concerned about reverse causality in my OLS regressions, as it could be the case that passive ownership happens to be higher for firms with earnings that are hard to predict in the first place. Column 5 shows that, reassuringly, there is essentially no relationship between earnings growth volatility and passive ownership. Finally, column 6, asks whether or not the absolute magnitude of earnings growth, normalized by the standard deviation of earnings growth over the past 8 quarters, is larger for high passive ownership stocks. As with column 5, there is no relationship between this quantity and passive ownership, again suggesting that high passive stocks do not have especially volatile earnings news.

	Num. Est.	St. Dev.	$\lvert Dist.\rvert$/P	$\lvert Dist.\rvert$/SD	SD(Earn.)	$\frac{\lvert\text{Growth}\rvert}{\text{SD(Earn)}}$
	(1)	(2)	(3)	(4)	(5)	(6)
Passive	-12.36***	0.00426	0.00874**	2.119***	-0.0208	-0.235
	(1.492)	(0.006)	(0.004)	(0.446)	(0.046)	(0.248)
Observations	225,073	225,073	225,073	225,073	210,017	195,093
R-squared	0.765	0.472	0.348	0.109	0.527	0.106
Weight	Equal	Equal	Equal	Equal	Equal	Equal
Controls	YES	YES	YES	YES	YES	YES
Fixed Effects	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ

Table D5 Cross-sectional regression of analyst coverage and earnings volatility on passive ownership. Table with estimates of β from:

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

where $\text{Outcome}_{i,t}$ is either the number of analysts (Num. Est.), the standard deviation of analyst estimates (St. Dev.), the absolute difference between actual earnings and the mean estimate of earnings ($\lvert Dist.\rvert$) normalized by the pre-earnings announcement price ($\lvert Dist.\rvert$/P), the absolute difference between actual earnings and the mean estimate of earnings, normalized by the standard deviation of analyst estimates ($\lvert Dist.\rvert$/SD), the standard deviation of year-over-year earnings growth over the past 8 quarter (SD(Earn)) and the absolute year-over-year earnings growth divided by the standard deviation of year-over-year earnings growth over the past 8 quarters $\left(\frac{\lvert\text{Growth}\rvert}{\text{SD(Earn)}}\right)$. 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.

Table D6 examines the relationship between investor attention, earnings news complexity, and passive ownership. One way to quantify attention 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 Bloomberg search index measures of Ben-Rephael et al. (2017) on passive ownership. The sample is stock/month observations between 2010 and 2019 that can be linked between Bloomberg and CRSP on ticker. All the controls and fixed effects are identical to Equation 4. Columns 1, 2 and 3 of Table D6 show that passive ownership is correlated with less institutional investor attention.

As an alternative way to measure investors' learning behavior, I examine downloads of SEC filings, with fewer downloads implying decreased gathering of fundamental information (Loughran and McDonald, 2017). Specifically, I examine 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. Column 4 of Table D6 shows that passive ownership is correlated with fewer downloads of SEC filings.

As discussed in the main body of the paper, another possible source of reverse causality would be if the earnings news of stocks with more passive ownership was generally harder to interpret. To test this, I obtained the complexity score from Loughran and McDonald (2020). This is a measure based on the text of 10-K filings, with higher values denoting more complexity. The second is the natural logarithm of the total size of the 10-K filing document, which Loughran and McDonald (2020) argue is also related to firm complexity.

	BBSI	High Attn.	BBC	ln(NR Total)	Complexity	File Size	CAV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Passive	-1.670***	-0.263***	-1.439***	-0.726***	0.290***	0.149	-12.21***
	(0.461)	(0.082)	(0.374)	(0.254)	(0.042)	(0.096)	(3.697)
Observations	64,831	64,831	64,831	517,876	81,966	81,966	408,471
R-squared	0.509	0.377	0.554	0.807	0.653	0.676	0.047

Table D6 Cross-sectional regression of attention and complexity on passive ownership. Table with estimates of β from:

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

where $\text{Outcome}_{i,t}$ is either the Bloomberg Search Index ($BBSI$), an indicator variable for high Bloomberg Attention (High Attn.) or the continuous abnormal institutional attention measure from Ben-Rephael et al. (2017) (BBC), one plus the natural logarithm of the number of non-robot downloads from Loughran and McDonald (2017) (ln(NR Total)), the Complexity and natural logarithm of 10-K document size from Loughran and McDonald (2020) (Complexity, ln(Doc. Size)) or cumulative abnormal pre-earnings announcement volume (CAV). 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. Includes earnings announcements between 1990 and 2019.

Column 5 shows that complexity is positively correlated with passive ownership in the cross-section. This suggests reverse causality may be an issue for the OLS regressions because e.g., if passive ownership is correlated with complexity, we might naturally expect such firms to have less informative prices. The effect, however, is economically small. Mean complexity is 0.392, so a 24% increase in passive ownership (i.e., moving from the 10th percentile to the 90th percentile in 2019) predicts a 0.07 increase in complexity i.e., less than 20% of its mean. Further, column 6 shows there is essentially no relationship between 10-K document size and passive ownership.

So far, these are just cross-sectional correlations, and do not speak to the only-through assumption underlying my IV design. Next, I examine how analyst coverage, earnings volatility, investor attention and complexity are affected by S&P 500 and Russell index changes. Rather than use an instrumental variables strategy, I directly examined the effects of index changes on these quantities with the following regression:

$$\text{Outcome}_{i,t} = \alpha + \beta_1 Post_{i,t} + \beta_2 Post_{i,t} \times Treated_{i,t} + FE + \epsilon_{i,t} \tag{D9}$$

where $Post_{i,t}$ is an indicator variable for the post-switching period and $Treated_{i,t}$ is an indicator variable for those that received the treatment of switching indices. FE are firm-by-cohort fixed effects (Coles et al., 2022). The coefficient of interest is β_2 i.e., the effect of switching indices on the outcome of interest for the treated group relative to the control group. These are essentially reduced form regressions in my IV setting, except I am omitting the $PassiveGap_{i,t}$ variable for easier interpretation.

Further, as a strong first stage is not required in this setting, I also consider stocks which switched from the Russell 2000 to the Russell 1000, as these are stocks moving to an index of larger firms and therefore may receive more analyst coverage/attention. I also consider deletions from the S&P 500 as these are stocks moving to indices of smaller firms (e.g., the MidCap or SmallCap) and therefore may receive less analyst coverage/attention.

Table D7 contains the results on analyst coverage and earnings volatility. The top panel presents the results for firms switching between the Russell 1000 and 2000. First, I'd like to highlight that none of the results are strongly statistically significant. That being said, the point estimates suggest that stocks switching into the Russell 1000 get an increase in coverage (both in terms of number of estimates, and a smaller difference between actual and expected earnings), while stocks switching to the Russell 2000 get a decrease in coverage. Reassuringly for my IV results in Table 3, for firms switching into the Russell 2000, index switching has essentially no effect on the volatility of earnings announcement news itself.

The second panel contains results for the S&P setting. Here, possibly because there are more than 5× as many observations as for Russell switchers, more of the estimated coefficients are statistically significant. Unsurprisingly, stocks added to the 500 receive an increase in the number of analysts covering the stock, while stocks dropped from the 500 receive a decrease in the number of analysts covering the stock. Interestingly, for both additions and deletions, analyst coverage becomes less accurate, and this effect is stronger for additions.

Another number that stands out from Table D7 is the increase in earnings volatility for firms added to the S&P 500. This, however, may be a consequence of the earnings-gaming some firms engage in to be added to the index. An example of this is Tesla, which was large enough to be added to the S&P 500, but did not meet S&P's profitability thresholds. One contribution to Tesla becoming profitable enough to be added to the S&P 500 was the sale of regulatory credits, which the company knew was not going to be a sustainable source of profit.

Table D8 contains the results on investor attention and complexity. Again, the top panel presents the results for firms switching between the Russell 1000 and the Russell 2000. And, like in Table D7, most of the estimated effects are not statistically significant. The point estimates provide suggestive evidence that stocks switching to the Russell 1000 receive an increase in attention. The evidence is more mixed for stocks switching to the Russell 2000, as they receive fewer downloads of SEC filings, but have a higher level of Bloomberg search attention. Finally, switching between Russell indices doesn't seem to be correlated with changes in complexity or 10-K filing size. This makes sense, as these are firm-level characteristics that are unlikely to change after a firm moves between totally market-capitalization-based indices.

The bottom panel presents results for S&P 500 index changes. Like with analyst coverage, being added to the 500 is correlated with increased attention, while being dropped is correlated with decreased attention, at least in terms of Bloomberg searches and downloads of SEC filings. As with the Russell index changes, there is little effect on complexity or 10-K document size. Firms added to the 500 have a slight increase in complexity, but a slight decrease in document size, so the evidence is mixed at best.

With all that being said, my takeaways from Tables D5, D6, D7 and D8 are as follows: First, in the cross-section, passive ownership is correlated with less analyst coverage in both quantity and quality. This raises the concern of reverse causality for the OLS regressions, and a possible violation of the only-through assumption for the IV regressions. That being said, the two index inclusion experiments have different effects on analyst coverage, with stocks switching to the Russell 2000 receiving less/worse coverage and stocks being added to the S&P 500 receiving more/better coverage.

This makes the consistency between the two experiments – in terms of the effect of passive ownership on pre-earnings announcement price informativeness – more surprising. In one experiment, investors seem to be paying more attention, while in the other they are paying less attention. In both cases, however, the instrumented change in passive ownership

	Panel A: Russell											
	2000 → 1000						1000 → 2000					
	Num. Est.	St. Dev.	$\|Dist.\|$/P	$\|Dist.\|$/SD	SD(Earn.)	$\frac{\|\text{Growth}\|}{\text{SD(Earn)}}$	Num. Est.	St. Dev.	$\|Dist.\|$/P	$\|Dist.\|$/SD	SD(Earn.)	$\frac{\|\text{Growth}\|}{\text{SD(Earn)}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post × Treated	0.2399	0.0008	-0.0003	-0.0824	0.0109*	-0.0161	-0.2926*	-0.0007	0.0003	0.0376	-0.0009	0.0821
	(0.1530)	(0.0010)	0.0000	(0.0750)	-0.006	-0.054	(0.1660)	(0.0010)	0.0000	(0.0720)	-0.008	-0.052
Post	1.6356***	-0.0014**	-0.0003	0.0963*	-0.0112***	0.1076***	0.1744	-0.0008	0.0010***	0.0812	-0.0099	-0.0297
	(0.1170)	(0.0010)	0.0000	(0.0490)	-0.004	-0.035	(0.1450)	(0.0010)	0.0000	(0.0550)	-0.007	-0.042
Observations	36,119	36,119	36,119	36,119	34848	33081	25155	25155	25155	25,155	24363	23226
R-squared	0.77	0.478	0.308	0.191	0.638	0.178	0.808	0.52	0.309	0.15	0.665	0.136
	Panel B: S&P 500											
	Additions						Deletions					
	Num. Est.	St. Dev.	$\|Dist.\|$/P	$\|Dist.\|$/SD	SD(Earn.)	$\frac{\|\text{Growth}\|}{\text{SD(Earn)}}$	Num. Est.	St. Dev.	$\|Dist.\|$/P	$\|Dist.\|$/SD	SD(Earn.)	$\frac{\|\text{Growth}\|}{\text{SD(Earn)}}$
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post × Treated	1.7159***	0.0038***	0.0002	-0.1785**	0.0219***	-0.0982*	-1.2604***	-0.0037	0.0008	-0.1056	0.0165	-0.2157*
	(0.1860)	(0.0010)	0.0000	(0.0800)	-0.007	-0.058	(0.4060)	(0.0020)	(0.0010)	(0.1660)	-0.015	-0.119
Post	1.5216***	-0.0023***	0.0001	-0.0489	-0.0182***	0.0636**	0.6519***	0.0010*	0.0009***	0.0843	-0.0051	0.0479
	(0.1100)	(0.0010)	0.0000	(0.0340)	-0.004	-0.026	(0.1450)	(0.0010)	0.0000	(0.0590)	-0.004	-0.039
Observations	191,752	191,752	191,752	191,752	184135	174747	90,794	90,794	90,794	90,794	88138	84905
R-squared	0.843	0.611	0.307	0.167	0.693	0.173	0.896	0.684	0.371	0.222	0.791	0.253

Table D7 Regression estimates for effect of index changes on analyst coverage and earnings volatility. Estimates from:

$$\text{Outcome}_{i,t} = \quad \alpha + \beta_1 Post_{i,t} + \beta_2 Post_{i,t} \times Treated_{i,t} + FE + \epsilon_{i,t}$$

where $\text{Outcome}_{i,t}$ is either the number of analysts (Num. Est.), the standard deviation of analyst estimates (St. Dev.), the absolute difference between actual earnings and the mean estimate of earnings ($|Dist.|$) normalized by the pre-earnings announcement price ($|Dist.|$/P), the absolute difference between actual earnings and the mean estimate of earnings, normalized by the standard deviation of analyst estimates ($|Dist.|$/SD), the standard deviation of year-over-year earnings growth over the past 8 quarter (SD(Earn)) and the absolute year-over-year earnings growth divided by the standard deviation of year-over-year earnings growth over the past 8 quarters $\left(\frac{|\text{Growth}|}{\text{SD(Earn)}}\right)$. Panel A contains observations from Russell rebalancing, while Panel B contains observations from S&P 500 index changes. FE are fixed effects for each cohort. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

						Panel A: Russell						
	2000 → 1000						1000 → 2000					
	BBC	High Attn.	ln(NR Total)	Complexity	ln(Doc. Size)	CAV	BBC	High Attn.	ln(NR Total)	Complexity	ln(Doc. Size)	CAV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post × Treated	0.0256	0.00	0.0523	0.00	0.0053	0.1	0.0516*	-0.0051	-0.0316	0.00	0.0153	0.05
	(0.025)	(0.005)	(0.036)	(0.005)	(0.031)	(0.234)	(0.028)	(0.014)	(0.030)	(0.008)	(0.020)	(0.231)
Post	0.0045	0.01	0.6005***	0.0291***	0.1032***	-0.65***	-0.2810***	-0.0281***	0.4809***	0.0363***	0.1118***	-0.49**
	(0.035)	(0.006)	(0.095)	(0.005)	(0.023)	(0.194)	(0.063)	(0.006)	(0.089)	(0.008)	(0.022)	(0.212)
Observations	5,462	5,462	48,620	8,595	8,595	47,588	1996	1996	28962	5,334	5,334	32,256
R-squared	0.511	0.315	0.618	0.73	0.646	0.065	0.455	0.25	0.639	0.776	0.677	0.057
						Panel B: S&P 500						
	Additions						Deletions					
	BBC	High Attn.	ln(NR Total)	Complexity	ln(Doc. Size)	CAV	BBC	High Attn.	ln(NR Total)	Complexity	ln(Doc. Size)	CAV
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Post × Treated	0.0401	0.0116**	0.2605***	0.0112*	-0.0308	-0.400**	-0.0999*	-0.0116	-0.2266***	-0.0117	-0.055	-0.095
	(0.025)	(0.005)	(0.035)	(0.006)	(0.019)	(0.164)	(0.057)	(0.012)	(0.066)	(0.016)	(0.095)	(0.366)
Post	-0.0429**	-0.0021	0.6216***	0.0392***	0.1440***	-0.345*	-0.1654***	-0.0258***	0.6047***	0.0341***	0.1268***	-0.565***
	(0.019)	(0.006)	(0.069)	(0.005)	(0.013)	(0.184)	(0.025)	(0.006)	(0.072)	(0.006)	(0.022)	(0.200)
Observations	97,876	97,876	197,310	46,962	46,962	283,900	40,526	40,526	129,617	19,696	19,696	116,016
R-squared	0.632	0.455	0.743	0.741	0.7	0.075	0.621	0.436	0.774	0.8	0.755	0.1

Table D8 Regression estimates for effect of index changes on investor attention and complexity. Estimates from:

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

where $\text{Outcome}_{i,t}$ is either the measure of continuous abnormal institutional attention from Ben-Rephael et al. (2017) built on Bloomberg searches (BBC), one plus the natural logarithm of the number of non-robot downloads from Loughran and McDonald (2017) (ln(NR Total)), the Complexity and natural logarithm of 10-K document size from Loughran and McDonald (2020) (Complexity, ln(Doc. Size)) or cumulative abnormal pre-earnings announcement volume (CAV). Panel A contains observations from Russell rebalancing, while Panel B contains observations from S&P 500 index changes. FE are fixed effects for each cohort. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

predicts decreases in pre-earnings announcement price informativeness. In fact, this calls into question whether or not analyst coverage and/or attention affect pre-earnings announcement price informativeness, and suggests that a violation of the only-through assumption is likely not driving my IV results.

In terms of the nature of earnings announcements themselves, there is also mixed evidence. While passive ownership is not correlated with earnings volatility in the cross-section, it is correlated with 10-K complexity. This is further evidence why the IV setting is crucial to rule out concerns of reverse causality. In the index switching setting, there is little relationship between index switching and complexity, allying concerns that reverse causality is driving these results. The exception to this is the strong relationship between being added to the S&P 500 and earnings volatility. But, as described above, this may be due to earnings gaming by firms wishing to be added to the index.

E Expanding the Sample

In the main body of the paper, I consider data between 1990 and 2019. In Table E9, I replicate the OLS regression results in Table 2 for different sets of years. The middle panel expands the sample to include 2020-2022, which is where we can see a significant difference from the results in the baseline sample (replicated in the top panel). The coefficient for $|Ret|$ shrinks by a little more than a $1/3^{rd}$, while the coefficients on $|Ret|/SD$ and PJ are nearly cut in half. The coefficient on IVD shrinks by about a third as well.

My proposed explanation for this is that in times of high volatility, the learning problem becomes harder, and generally weakens the relationship between passive ownership and pre-earnings announcement price informativeness. In what follows, I present several pieces of evidence consistent with this explanation.

To start, the bottom right panel of Table E9 examines 2020-2022 on its own, with the aim of understanding why including those years seems to have such a dramatic effect on the results. The estimated coefficients for $|Ret|$ and $|Ret|/SD|$ are significantly smaller than those in the original sample, while the coefficient for PJ increases and the coefficient on IVD switches sign.

It seems counterintuitive that in the PJ regression the coefficient on passive ownership is larger in the 2020-2022 sample than in the original sample, yet the coefficient is smaller than the original estimate when using all years between 1990 and 2022. I want to highlight, however, that it is not straightforward to directly compare the results in the bottom panel to those in the other panels. This is because the firm fixed effects and coefficients on the control variables can be substantially different. For example, in the PJ regression, when using the 1990-2019 sample, the coefficient on institutional ownership is 0.04 and highly statistically significant (t-statistic of 3.31) while in the 2020-2022 sample, the coefficient is -0.011 and insignificant (t statistic of -0.15). In addition, the means and standard deviations of PJ are about 25% lower in the 2020-2022 sample than the 1990-2019 sample, which again makes directly comparing the magnitudes of the estimated effects difficult.

The natural next question is what happens to the IV estimates when expanding the sample to 2022. In Table E10, I replicate the results in Table 3 including data from 2020-2022. Like with the OLS regression estimates, the IV estimates shrink in size when including these extra years, but they maintain their statistical significance. The exception to this is IVD for the S&P experiment. But, as highlighted in Table E9, the relationship between IVD and passive ownership seems to have been the most affected – relative to the other measures of pre-earnings announcement price informativeness – during the COVID years.

Tables E9 and E10 suggest the relationship between passive ownership and pre-earnings announcement price informativeness were dramatically different in the 2020-2022 period, relative to the 1990-2019 period. My proposed explanation is built on the fact that the COVID crisis was one of the biggest spikes in realized volatility in the history of the US stock market (Baker et al., 2020).

The logic for this affecting my results is as follows: for $|Ret|/SD$, PJ and IVD, there is a normalization of what happened on the earnings announcement itself based on volatility around the announcement. And because overall volatility increased, and was especially high around non-earnings-related information events – e.g., news about monetary policy, COVID cases and lockdowns – the stock market response to earnings information might seem relatively less important.

Recall, for example, how IVD is defined – it is the difference in implied volatility between options that span earnings announcements and those that are less exposed to earnings announcement risk. If overall volatility was higher for reasons unrelated to earnings announcements, IVD should decline. Consistent with this, Figure E.11 shows the equal-weighted cross-sectional averages of the pre-earnings announcement price informativeness measures

	Panel A: 1990-2019			
	$\|Ret\|$	$\|Ret\|/SD$	PJ	IVD
	(1)	(2)	(3)	(4)
Passive	0.0635***	4.425***	0.248***	0.112***
	(0.007)	(0.384)	(0.060)	(0.030)
Observations	441,238	441,238	148,864	111,604
R-squared	0.194	0.214	0.163	0.3
	Panel B: 1990-2022			
	$\|Ret\|$	$\|Ret\|/SD$	PJ	IVD
	(5)	(6)	(7)	(8)
Passive	0.0386***	2.531***	0.127**	0.0714***
	(0.008)	(0.485)	(0.055)	(0.025)
Observations	476,932	476,932	161,859	133,048
R-squared	0.192	0.207	0.159	0.274
	Panel C: 2020-2022			
	$\|Ret\|$	$\|Ret\|/SD$	PJ	IVD
	(9)	(10)	(11)	(12)
Passive	0.00431	1.095	0.607**	-0.0183
	(0.022)	(0.985)	(0.253)	(0.079)
Observations	35,586	35,586	12,620	21,230
R-squared	0.249	0.235	0.31	0.323
Weight	Equal	Equal	Equal	Equal
Controls	YES	YES	YES	YES
Fixed Effects	Firm, YQ	Firm, YQ	Firm, YQ	Firm, YQ

Table E9 Cross-sectional regression of price informativeness on passive ownership for different sets of years. Table with estimates of β from:

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

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$. 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.

	Panel A: Russell				
	First Stage	Instrumental Variables			
	Passive	$\|Ret\|$	$\|Ret\|/SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)
Post × Treated	1.48***				
× Passive Gap	(0.119)				
Post	0.02***				
	(0.002)				
Passive		0.15***	5.88**	0.64*	0.23***
		(0.036)	(2.257)	(0.379)	(0.082)
Observations	34,332	34,332	34,332	13,003	12,636
F-Statistic	251.6				

	Panel B: S&P 500				
	First Stage	Instrumental Variables			
	Passive	$\|Ret\|$	$\|Ret\|/SD$	PJ	IVD
	(1)	(2)	(3)	(4)	(5)
Post × Treated	0.407***				
× Passive Gap	(0.061)				
Post	0.022***				
	(0.002)				
Passive		0.180***	7.130***	0.930***	0.093
		(0.044)	(1.872)	(0.282)	(0.141)
Observations	291,199	291,199	291,199	109,746	167,063
F-Statistic	201.1				

Table E10 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness, 1990-2022. Estimates from:

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

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

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$ 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 2-4 are instrumental variables 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. Includes data from 1990-2022.

from 1990-2022. While all the measures were trending steadily upward from the start of my sample in 1990 to the end of my original sample in 2019 Q4 (vertical red line), all of them have a significant drop at the start of the COVID-19 pandemic.

At face value, this logic of differences in unconditional volatility driving the changes to my results in 2020-2022 doesn't seem to directly apply to $|Ret|$, which is not normalized by unconditional volatility. Despite this, Figure E.11 shows that $|Ret|$ also dipped right after the onset of the COVID pandemic. As I explain below, I believe this is partially due to a decreased emphasis on earnings news during the pandemic, as such news was possibly less salient than the risk associated with e.g., COVID cases and lockdowns.

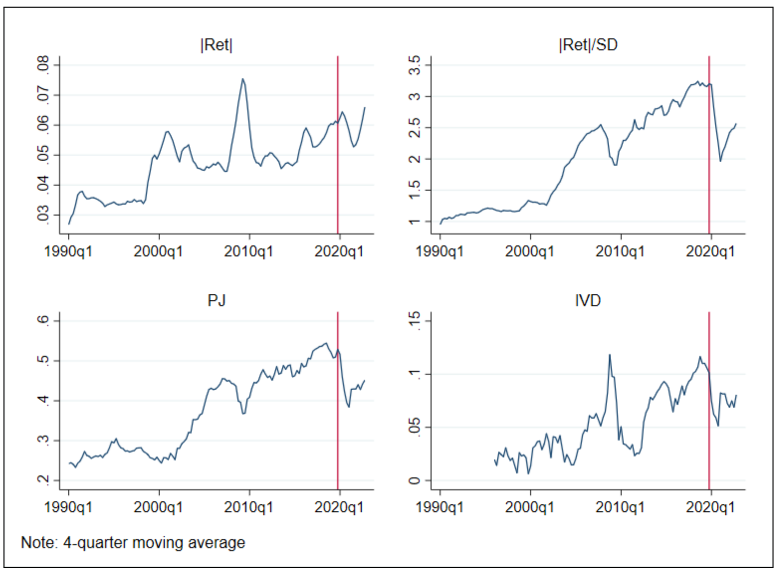

Figure E.11. Trends in $|Ret|$, $|Ret|/SD$, PJ and IVD, 1990-2022. Equal-weighted average of $|Ret|$, $|Ret|/SD$, PJ and IVD each quarter. 4 quarter moving average filter is applied after taking the average each quarter. Vertical line at 2019 Q4.

All that being said, if there was just a level shift down in pre-earnings announcement price informativeness, it's not obvious why my results would be weaker when including the 2020-2022 data. All my regression specifications include time fixed effects, which should take care of this common time-series component.

My proposed explanation for why the relationship between passive ownership and pre-earnings announcement price informativeness was weaker from 2020 to 2022 than in the earlier sample is that during periods of high volatility, the learning problem becomes harder.

The logic is that, in an Admati (1985)-style model, if fundamental volatility increases, for a given amount of signal precision, price informativeness would decline.

To test this hypothesis, I need to construct a measure of aggregate stock market volatility. To this end, I replicated and expanded the results in Campbell et al. (2001), specifically the measure of aggregate idiosyncratic volatility, which I plot in Figure E.12. During the matched sample, my results closely mirror those in the original paper, suggesting a successful replication.

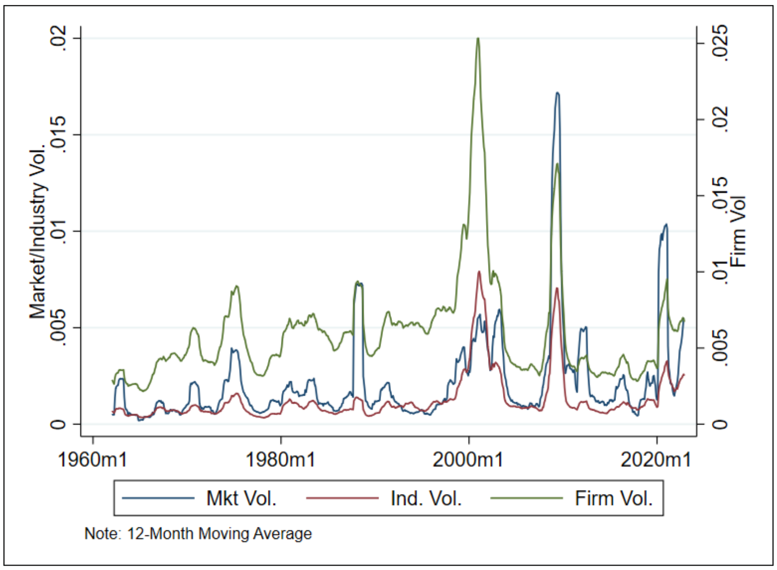

Figure E.12. Replication and extension of Campbell et al. (2001) Decomposition of aggregate volatility into a market, industry and firm-specific components following Campbell et al. (2001).

Next, I augment my baseline regression, including aggregate firm volatility from Campbell et al. (2001), as well as an interaction term between aggregate firm volatility and passive ownership. The volatility measure does not vary at the firm-level (only at the month-level), so I switch to year fixed effects instead of year-quarter fixed effects. This implies that the measure of volatility is exploiting within-year differences in aggregate idiosyncratic volatility. This regression specification is useful, as it will show whether the weaker relationship between passive ownership and price informativeness was a phenomenon specific to the COVID pandemic, or a more general feature of periods with increased volatility.

The results are in Table E11. The first 4 columns are a replication of the OLS regression

	$\lvert Ret \rvert$ (1)	$\lvert Ret \rvert /SD$ (2)	PJ (3)	IVD (4)	$\lvert Ret \rvert$ (5)	$\lvert Ret \rvert /SD$ (6)	PJ (7)	IVD (8)
Passive	0.0399***	2.711***	0.133**	0.0763***	0.0461***	4.534***	0.262***	0.115***
	(0.008)	(0.481)	(0.055)	(0.027)	(0.010)	(0.558)	(0.083)	(0.037)
Firm Vol					1.090***	0.308	-1.396*	3.824***
					(0.152)	(4.473)	(0.767)	(0.599)
Firm Vol × Passive					-0.876	-332.4***	-24.04**	-7.388
					(1.569)	(64.920)	(10.040)	(4.710)
Observations	476,932	476,932	161,859	133,048	476,932	476,932	161,859	133,048
R-squared	0.187	0.202	0.156	0.241	0.19	0.203	0.156	0.25
Weight	Equal	Equal	Equal	Equal	Equal	Equal	Equal	Equal
Controls	YES	YES	YES	YES	YES	YES	YES	YES
Fixed Effects	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year	Firm, Year

Table E11 Cross-sectional regression of price informativeness on passive ownership. Table with estimates of β from:

$$\text{Price informativeness}_{i,t} = \alpha + \beta_1 Passive_{i,t} + \beta_2 FirmVol_t + \beta_3 Passive_{i,t} \times FirmVol_t + \gamma X_{i,t} + \phi_\tau + \psi_i + e_{i,t}$$

where Price informativeness$_{i,t}$ is either $|Ret_{i,t}|$, $|Ret|/SD_{i,t}$, $PJ_{i,t}$ or $IVD_{i,t}$. 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. $FirmVol_t$ is the measure of aggregate firm volatility in month t from Campbell et al. (2001). Standard errors double clustered at the firm and year-quarter level in parenthesis. Sample is data between 1990 and 2022.

results in Table E9 for the 1990-2022 sample. The next 4 columns include aggregate firm volatility on the right-hand-side, as well as the interaction term between aggregate firm volatility and passive ownership. For all 4 measures, the point estimate on passive ownership increases in columns 5-8 relative to columns 1-4. This effect is economically large, with the coefficient increasing by 16% for $|Ret|$, 67% for $|Ret|/SD$, 97% for PJ and 51% for IVD. Further, the coefficient on the interaction term is strongly negative. This suggests that, consistent with my proposed mechanism, the relationship between passive ownership and price informativeness is weaker during periods of elevated idiosyncratic volatility.

Finally, I want to highlight that this is not the only reason why the relationship between passive ownership and pre-earnings announcement price informativeness might be weaker during periods of elevated volatility. For example, it could be that during periods of high volatility, investors focus their limited attention on systematic risk, instead of idiosyncratic risk (see e.g., Kacperczyk et al. (2016)). The logic is that if all investors stop learning about firm-specific risk, the gap in pre-earnings announcement price informativeness between stocks with high and low passive ownership should shrink.