Online Appendix for Passive Ownership and Price Informativeness

MARCO SAMMON*

June, 2022

A Robustness of Stylized Facts

A.1 Decline of pre-earnings drift by SUE quintile

In figure 1, it appears as though the time series trend in pre-earnings drift was different between firms that ended up releasing good news, and firms that ended up releasing bad news. To quantify this, figure A.1 plots the equal-weighted average of the pre-earnings drift magnitude, $DM_{i,t}$, within quintiles of SUE, formed each quarter. While the average pre-earnings drift has been declining for all 5 quintiles, the asymmetry initially arose in 2010 and has been growing since then – with a larger decline in pre-earnings drift for firms that ended up releasing bad news.

In section C.4 below, I show this asymmetry is stronger among stocks with more passive ownership and is not solely explained by the fact that passive owners may be more willing to lend out their shares than other institutional investors.

A.2 Decomposition of earnings days' share of volatility

Figure A.2 decomposes the rise of $r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2$ i.e., the decline of QVS into rise in the numerator (volatility on earnings days) and the denominator (volatility on leading up to the

*Harvard Business School. Email: mcsammon@gmail.com.

Figure A.1. Average pre-earnings drift by quintile of SUE. Each quarter, I sort firms into 5 buckets based on SUE, and compute the equal-weighted average DM within each bucket. This figure plots the 8-quarter moving average of this quantity within each quintile of SUE.

earnings announcement plus volatility on the earnings day itself). The trend in QVS was driven by a simultaneous increase in the numerator, and decrease in the denominator.

A.3 Relationship between DM and QVS

As mentioned in the main body of the text, while DM and QVS are related, they can capture different information. First, a univariate regression of QVS on DM has an R-squared of just under 50%, so they are far from perfectly correlated. To understand where they differ, figure A.3 presents a scatter plot with QVS on the y-axis and DM on the x-axis. This shows that if DM is low, QVS will also tend to be low, while if DM is high, QVS can take on any value. As discussed in the main body of the paper, DM and QVS will tend to disagree when the earnings-day return is large – leading to low values of DM – but, there was significant volatility in the pre-earnings announcement period – leading to relatively higher values of QVS.

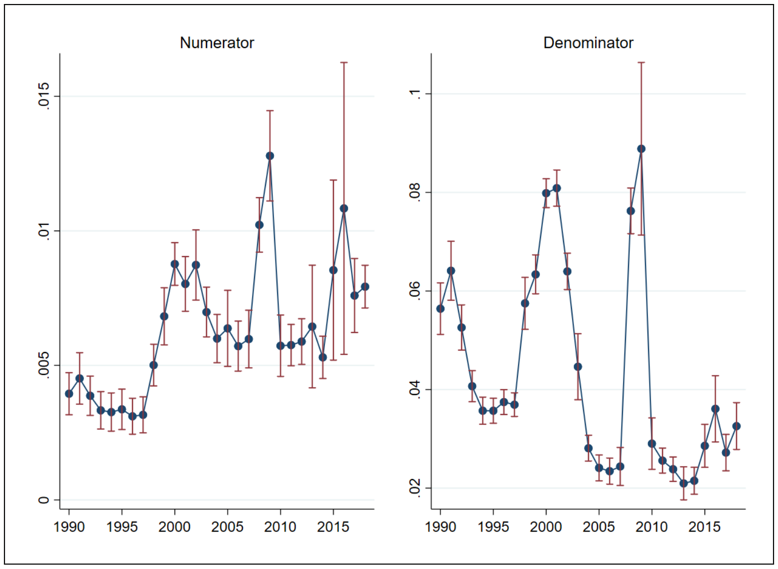

Figure A.2. Decomposition of QVS. This figure plots coefficients from a regression of the numerator and denominator of the term in parenthesis in QVS on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient. For firm i around earnings announcement t the quadratic variation share (QVS) is defined as: $QVS_{i,t} = 1 - \left(r_{i,t}^2 / \sum_{\tau=-22}^{0} r_{i,t+\tau}^2 \right)$, where r denotes a market-adjusted daily return. The numerator of the term in parenthesis is the squared earnings-day return, while the denominator is the sum of squared returns from $t-22$ to t. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.

A.4 Placebo tests for stylized facts

In this subsection, I conduct three placebo tests for the stylized facts: (1) Using the start of the pre-earnings announcement window as a placebo announcement date (2) Using randomly selected days and (3) Using FOMC meetings as placebo announcement dates. All three tests show that my results are specific to earnings-announcement days.

A.4.1 22 Days before each earnings announcement

For all the placebo tests, I run the following regression:

$$Outcome_{i,t} = \alpha + \sum_{k=1990}^{2018} 1_{year(t)=k} + \phi_q + \epsilon_{i,t} \tag{1}$$

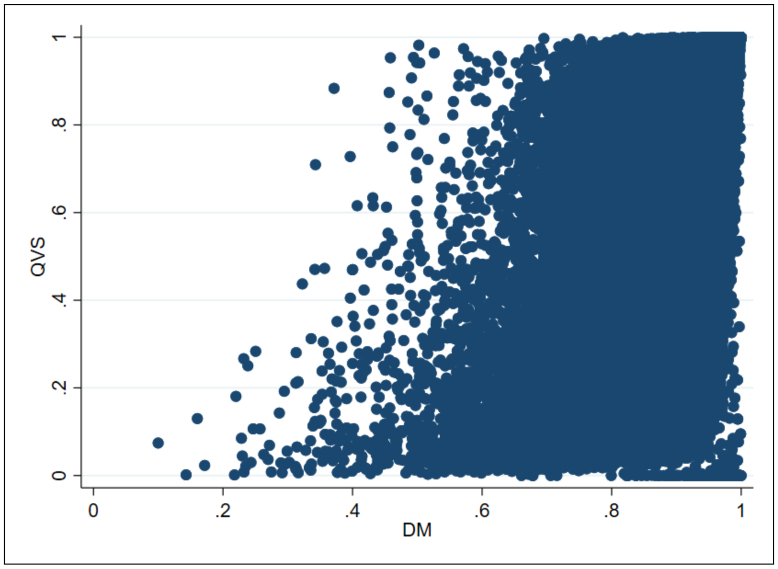

Figure A.3. QVS vs. DM. This figure is a scatter plot of QVS on DM. Each blue dot represents a single earnings announcement for a single firm.

Here, $Outcome$ is DM or QVS computed with respect to the placebo earnings announcement dates, which in this case are 22 trading days before the actual announcements. $\sum_{k=1990}^{2018} 1_{year(t)=k}$ are a set of dummy variables for each year between 1990 and 2018. ϕ_q is a quarter-of-year fixed effect to account for seasonal patterns. Standard errors are clustered at the security level.

I select these days because they are the start of the pre-earnings window for both price informativeness measures. Figure A.4 shows that there is no trend toward decreased price informativeness before these placebo earnings dates.

A.5 Random days

In this sub-section, I randomly select 4 days each year to be placebo earnings announcements, and re-run regression 1. Although this seems like a more natural test than using the dates 22 trading days before each earnings announcement, it has the disadvantage that there can be overlapping placebo announcements in the 23-day windows of interest. Figure A.5 shows that for both measures, there is no trend toward decreased informativeness before these placebo earnings dates either. In unreported results, I try several different seeds

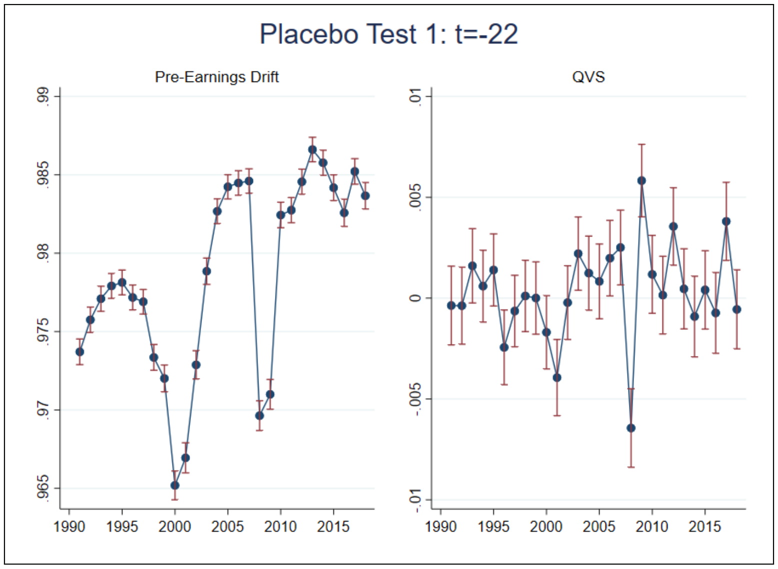

Figure A.4. Placebo Test 1: 22 trading days before each earnings announcement. Each panel presents coefficients from a regression of each of the price informativeness measures on a set of year fixed effects. These regressions also include quarter-of-year fixed effects to account for seasonality. The red bars represent 95% confidence intervals. Standard errors are clustered at the security level.

for identifying the random placebo earnings dates, and find no qualitative or quantitative difference from Figure A.5.

A.5.1 Scheduled FOMC Announcement Dates

The final set of placebo earnings days are FOMC announcements. To create an apples-to-apples comparison with the anticipated nature of earnings announcements, I restrict to scheduled FOMC meetings.

Figure A.6 plots coefficients on the year fixed effects from regression 1. There is no drop in average pre-FOMC drift. There is a slight trend toward increased volatility on FOMC announcement dates i.e., a trend toward lower QVS, but the magnitude is significantly smaller than the downward trend for actual earnings announcement dates documented in the main body of the paper (-0.25% vs. nearly -20%).

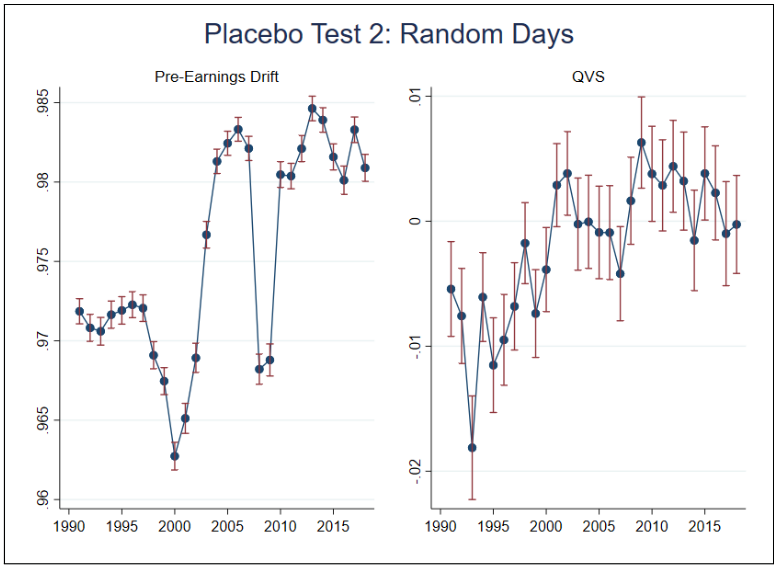

Figure A.5. Placebo Test 2: Randomly selected dates. Each panel presents coefficients from a regression of each of the price informativeness measures on a set of year fixed effects. These regressions also include quarter-of-year fixed effects to account for seasonality. The red bars represent 95% confidence intervals. Standard errors are clustered at the security level.

B Data details

B.1 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 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-variation in missing observations is not driving my results for two reasons: (1) I re-run every regression using only post-2000 data when ruling out the influence of Regulation Fair Disclosure and the results are similar and (2) These missing earnings times could only move the day I identify as the effective earnings date *earlier* in calendar time, which would bias both the pre-earnings drift and earnings-day volatility measures toward finding nothing.

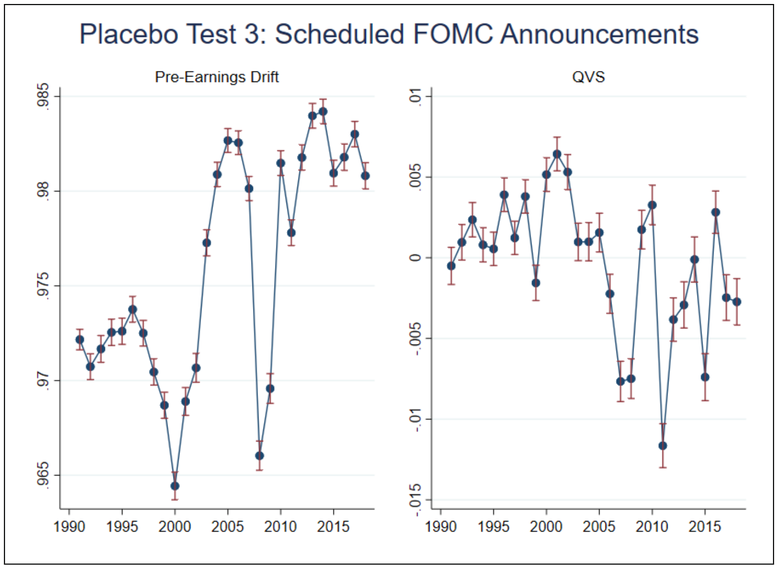

Figure A.6. Placebo Test 3: Scheduled FOMC announcement dates. Each panel presents coefficients from a regression of each of the price informativeness measures on a set of year fixed effects. These regressions also include quarter-of-year fixed effects to account for seasonality. The red bars represent 95% confidence intervals. Standard errors are clustered at the security level.

Specifically, it would lead to selecting days where no news was released, which likely have smaller, rather than larger moves on average, pushing DM toward 1, and QVS toward 4.3%.

B.2 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.[1] I use the WRDS MF LINKS database to connect the funds identified as passive in CRSP with the S12 data. If a security never appears in the S12 data, I assume its passive ownership is zero. S12 data is only reported at the end of each calendar quarter, so to get a monthly estimate of passive ownership, I linearly interpolate passive ownership between quarter-ends. All results are quantitatively unchanged if I instead fix passive ownership at its last reported level between the ends of calendar quarters.

Institutional ownership computed by adding up the holdings of all 13-F filings institutions to the CUSIP/Quarter level. It is then merged to CRSP on historical CUSIP or CUSIP. If a

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

CUSIP never appears in the 13-F data, institutional ownership is assumed to be zero. 13-F forms are filed quarterly, so I linearly interpolate institutional ownership between quarter-ends

B.3 CRSP volume vs. total volume

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

C Robustness of Cross-Sectional Regression Results

C.1 Relationship to Weller's price jump measure

Weller (2018) examines the effect of algorithmic trading activity on price informativeness. The key measure of price informativeness is the price-jump, defined as:

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

In words the price jump measure is: (Cumulative return from the day before the earnings announcement to b days after)/(Cumulative return from a days before the earnings announcement to b days after). Although this is not identical to any of my measures, it is related to both DM and QVS. In this subsection I (1) show that my results are robust to instead using his measure of price informativeness and (2) argue why my measures have some unique advantages in my setting.

Because this measure uses net, rather than gross returns, one might be concerned that $jump_{i,t}^{(a,b)}$ could be arbitrarily large, because $CAR_{i,t}^{(T-a,T+b)}$ can be close to zero. To solve this issue, Weller applies the filter described in equation 3 of his paper, which drops observations with absolute values of $CAR_{i,t}^{(T-a,T+b)}$ below a specific threshold. As a result, the price jump

	(1)	(2)	(3)
Passive Ownership	0.519***	0.474***	0.314***
	(0.058)	(0.061)	(0.064)
Observations	162,055	148,931	148,931
R-Squared	0.159	0.161	0.162
Firm + Year/Quarter FE	✓	✓	✓
Matched to Controls		✓	✓
Firm-Level Controls			✓
Weight	Equal	Equal	Equal

Table C.1Passive ownership and Weller's price jump measure. Estimates of β from:
$jump_{i,t}^{(22,2)} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
Where $jump_{i,t}^{(22,2)}$ is the price jump measure from Weller (2018) with $a = 22$ and $b = 2$. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

can only be computed for 45.5% of observations. My measures have the advantage that I don't need to apply such a filter, and they can be computed for every earnings event.

To ensure, however, that my results are not specific to the way I defined price informativeness, I re-run my baseline regressions with $jump_{i,t}^{(a,b)}$ on the left-hand-side, choosing $b = 2$ and $a = 22$. I choose $a = 22$ instead of Weller's original $a = 21$ to be consistent with the pre-earnings window in DM and QVS. For these results, I don't report the value-weighted regressions because the filter described above drops over half the observations, and this attrition is not evenly spread over the firm size distribution.

Table C.1 contains the results. Across all specifications, higher levels of passive ownership are correlated with higher price jump ratios. Consistent with my original results, this implies that passive ownership leads to less informative prices before earnings announcements.

C.2 Relationship to Manela's CAR and CATO measures

Manela (2014) examines price informativeness around FDA drug approvals. The two measures he uses to quantify price informativeness are cumulative abnormal returns (CAR) and cumulative abnormal turnover (CATO). In this subsection I (1) show that my results are robust to instead using his measure of price informativeness and (2) argue why my measures have some unique advantages in my setting.

One exercise in the paper is comparing the pre vs. day-of vs. post announcement cumulative abnormal returns (CAR). By looking at drug approvals, Manela's empirical exercises are focused on good news. Earnings announcements, however, may contain good or bad news. To create an analogue of Manela's pre vs. day-of vs. post announcement CAR results, I need to condition on the news itself. To this end, I split firms into 10 deciles based on their standardized unexpected earnings (SUE), and focus on firms in the top decile. I calculate the average pre-announcement ($t = -5$ to $t = -1$), announcement-day ($t = 0$ to $t = 1$) and post-announcement ($t = 2$ to $t = 6$) cumulative market-adjusted return (R_1, R_2 and R_3 in Table 2 of Manela (2014)).

I re-run my baseline regressions, using these various measures of CAR as the left-hand-side variables. I don't report the value-weighted regressions because the top 10% of earnings news is not equally spread out among the firm size distribution. Table C.2 contains the results. The table shows that high passive stocks who experience good news don't tend to have larger pre-earnings or post-earnings returns. They do, however, systematically tend to have larger earnings-day returns, consistent with the DM and QVS results in the main body of the paper.

In the mechanisms section of my paper, I examine pre-earnings announcement abnormal turnover. Manela does something similar, studying abnormal turnover around FDA drug approval events. He quantifies this using CATO, which is similar to my measure of cumulative abnormal turnover: it is defined as cumulative turnover for the stock minus the value-weighted average turnover in the market over the same period. The proposed interpretation is similar to my cumulative abnormal turnover (CAT) measure, with higher values of CATO implying more information being traded into prices.

To this end, I again split firms into 10 deciles based on their standardized unexpected earnings (SUE) and focus on firms in the top decile. I then calculate the average pre-announcement ($t = -5$ to $t = -1$), announcement-day ($t = 0$ to $t = 1$) and post-announcement ($t = 2$ to $t = 6$) CATO (T_1, T_2 and T_3 in Table 2 of Manela (2014)). Table C.3 contains the results. Consistent with what I found using CAT, the first column shows that passive ownership is correlated with less pre-earnings trading. Column 2 shows there

	R_1	R_2	R_3
Passive Ownership	0.0326	0.0706**	-0.0359
	(0.021)	(0.031)	(0.022)
Observations	31,571	31,571	31,571
R-Squared	0.262	0.244	0.243
Firm + Year/Quarter FE	✓	✓	✓
Matched to Controls	✓	✓	✓
Firm-Level Controls	✓	✓	✓
Weight	Equal	Equal	Equal

Table C.2Passive ownership and Manela's CAR. Estimates of β from:
$R_{j,i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
Where $R_{1,i,t}$ is the cumulative market-adjusted return from $t = -5$ to $t = -1$, $R_{2,i,t}$ is the cumulative market-adjusted return from $t = 0$ to $t = 1$ and $R_{3,i,t}$ is the cumulative market-adjusted return from $t = 2$ to $t = 6$. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

	T_1	T_2	T_3
Passive Ownership	-0.0389**	-0.0254	-0.0362*
	(0.018)	(0.019)	(0.019)
Observations	31,571	31,571	31,571
R-Squared	0.619	0.584	0.588
Firm + Year/Quarter FE	✓	✓	✓
Matched to Controls	✓	✓	✓
Firm-Level Controls	✓	✓	✓
Weight	Equal	Equal	Equal

Table C.3Passive ownership and Manela's CATO. Estimates of β from:
$T_{j,i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
Where $T_{1,i,t}$ is the cumulative abnormal turnover from $t = -5$ to $t = -1$, $T_{2,i,t}$ is the cumulative abnormal turnover from $t = 0$ to $t = 1$ and $T_{3,i,t}$ is the cumulative abnormal turnover from $t = 2$ to $t = 6$. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

is no statistically significant difference in average trading volume on the earnings day itself between high and low passive stocks. Finally, column 3 shows that high passive stocks have lower trading after the announcement, although this result is only marginally significant.

I believe CAT has some advantages relative to CATO in my setting. Specifically, one might be concerned e.g., that if passive owners don't trade, the presence of passive ownership would decrease average trading volume. Using my CAT measure, however, alleviates this concern because high passive ownership stocks' lower average trading would be incorporated into the denominator. So, even if passive ownership leads to less trading on average, it's not clear to me why – outside of my proposed mechanism – this lower trading would be concentrated in the period before earnings announcements.

C.3 Placebo tests for cross-sectional regressions

Table C.4 contains placebo tests for the cross-sectional regressions of QVS on passive ownership. For placebo earnings announcements, Panel A uses the date 22 trading-days before each earnings announcement, Panel B uses random days and Panel C uses scheduled FOMC announcements. During my sample (1990-2018), there are at least 22 trading days between each scheduled FOMC announcement, so there are no overlapping announcements in any of the 23-day windows (22 pre-placebo earnings days + the placebo earnings day itself). The row labeled "Cross-Sectional Estimates" is copied from the corresponding table in the main body of the paper.

In Panels A and B, the coefficients are all economically small and statistically insignificant. For Panel C, all the equal weighted coefficients are negative and statistically significant. This is evidence that prices may have become less informative before the release of systematic information. The magnitude of these coefficients, however, is small, at about 1/20th the size of the baseline cross-sectional estimates.

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

Figure A.1 shows that the decline of the pre-earnings drift differed depending on whether or not the firm released good or bad news. To quantify the role of passive ownership, I re-created the baseline pre-earnings drift regressions, but each quarter, I break earnings announcements into 5 quintiles based on SUE. I include dummy variables for each SUE quintile, as well as interaction terms between these dummy variables and passive ownership. For the regressions in this section, the middle quintile always is the omitted dummy variable.

Table C.5 contains the results. The last row of the table contains the p-value from a test of whether the interaction term between $Passive$ and $LowSUE$ is equal to the interaction term between $Passive$ and $HighSUE$. Across most specifications, the effects of passive ownership on the pre-earnings drift tend to be stronger for firms that end up releasing negative news than positive news. In the paper, I argue this is because overpricing is generally harder to correct than underpricing, as there are additional frictions associated with short selling relative to just buying a stock. In section E.4 below, I argue this asymmetry is not entirely due to the fact that passive owners are willing to lend out their shares, making it easier to short the underlying stocks.

	Panel A: 22 trading-days before each earnings announcement				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	0.00341	0.00258	0.00308	0.00521	0.000
	(0.007)	(0.007)	(0.007)	(0.025)	(0.028)
Observations	416,166	386,668	386,668	386,668	386,668
R-Squared	0.035	0.036	0.036	0.036	0.036
	Panel B: randomly selected dates				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	0.00158	-0.00134	-0.00571	0.0327	0.026
	(0.008)	(0.008)	(0.008)	(0.039)	(0.041)
Observations	386,327	352,500	352,500	352,500	353,546
R-Squared	0.036	0.037	0.037	0.031	0.032
	Panel C: scheduled FOMC announcements				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.0209***	-0.0210***	-0.0162**	-0.0255	-0.026
	(0.007)	(0.007)	(0.007)	(0.036)	(0.026)
Observations	985,513	902,595	902,595	902,595	902,595
R-Squared	0.025	0.026	0.026	0.031	0.032
Baseline Estimates	-0.524***	-0.501***	-0.408***	-0.214*	-0.232**
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.4 Cross-Sectional Regression Placebo Tests: Earnings-Day Volatility. Table with estimates of β from:

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

$QVS_{i,t} = 1 - \left(r^2_{i,t} / \sum_{\tau=-22}^{0} r^2_{i,t+\tau} \right)$, where t is a placebo earnings date. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

	(1)	(2)	(3)	(4)
Passive	-0.0302***	-0.0272***	-0.0489***	-0.0496***
	(0.008)	(0.007)	(0.016)	(0.012)
Low SUE	-0.00295***	-0.00265***	-0.000148	-8.00E-05
	(0.000)	(0.000)	(0.001)	(0.001)
2	-0.000773**	-0.000721**	-0.000569	-0.000508
	(0.000)	(0.000)	(0.001)	(0.001)
3	-0.00187***	-0.00204***	0.000416	-0.000231
	(0.000)	(0.000)	(0.001)	(0.001)
High SUE	-0.00418***	-0.00439***	0.00120*	0.000327
	(0.000)	(0.000)	(0.001)	(0.001)
Low SUE x Passive	-0.0341***	-0.0262***	-0.0239**	-0.0163
	(0.006)	(0.006)	(0.011)	(0.010)
2 x Passive	-0.00804*	-0.00386	-0.0155	-0.00673
	(0.005)	(0.005)	(0.011)	(0.010)
3 x Passive	0.0119**	0.00931*	0.013	0.0153*
	(0.005)	(0.005)	(0.009)	(0.008)
High SUE x Passive	0.0187***	0.00935*	0.00415	0.00134
	(0.006)	(0.005)	(0.009)	(0.008)
Observations	346,924	334,054	346,924	334,054
R-squared	0.215	0.235	0.255	0.274
Firm + Year/Quarter FE	✓	✓	✓	✓
Matched to Controls		✓		✓
Firm-Level Controls		✓		✓
Weight	Equal	Equal	Value	Value
p-Value	0.000	0.000	0.005	0.029

Table C.5Effect of passive ownership on the pre-earnings drift by quintile of SUE. Estimates of β, b_j and c_j from:

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

Where $DM_{i,t}$ is a measure of the pre-earnings drift. $1_{SUE_{i,t} \in Q_j}$ is an indicator for firm i being in the jth quintile of SUE at time t, where the quintiles are formed each quarter. For every regression, the middle quintile is the omitted group. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

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

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

$$DM_{it}^{n} = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t+n)}} & \text{if } r_{(t,t+n)} > 0 \\ \frac{1+r_{(t-22,t+n)}}{1+r_{(t-22,t-1)}} & \text{if } r_{(t,t+n)} < 0 \end{cases} \tag{2}$$

Figure C.1 shows that the time-series trends in DM are robust to choices of n up to 5. Table C.6 shows that including up to 5 days does not quantitatively or qualitatively change my baseline results.

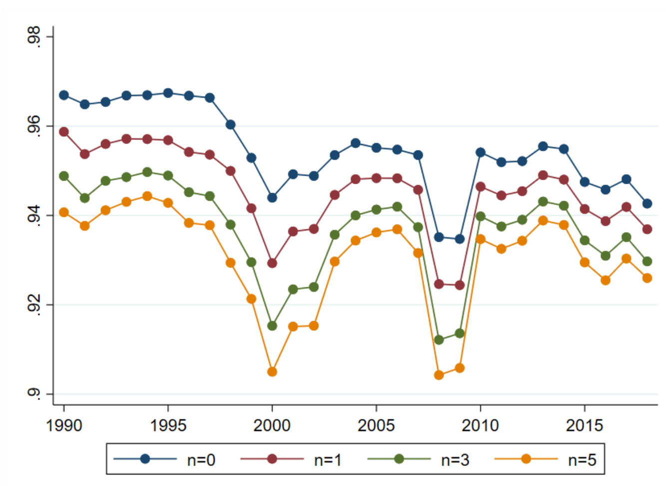

Figure C.1. Time series trends in DM^n. This figure plots coefficients from a regression of DM^n on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient.

The same concern could also apply to my results on QVS. In a similar vein, define an alternative measure which includes up to n days after t in the volatility attributed to the announcement itself:

$$QVS_{i,t}^{n} = 1 - \left[\left(\sum_{j=0}^{n} r_{i,t+j}^{2} \right) / \left(\sum_{\tau=-22}^{n} r_{i,t+\tau}^{2} \right) \right] \tag{3}$$

	Include t+1				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.0326***	-0.0380***	-0.0403***	-0.0475***	-0.0409***
	(0.007)	(0.007)	(0.007)	(0.016)	(0.012)
Observations	491,984	448,604	448,604	448,604	448,604
R-Squared	0.195	0.199	0.215	0.242	0.262
	Include t+1 to t+3				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.0147**	-0.0241***	-0.0341***	-0.0498***	-0.0500***
	(0.007)	(0.007)	(0.007)	(0.016)	(0.011)
Observations	491,925	448,569	448,569	448,569	448,569
R-Squared	0.198	0.201	0.218	0.229	0.249
	Include t+1 to t+5				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.00223	-0.0128	-0.0294***	-0.0421**	-0.0469***
	(0.008)	(0.008)	(0.008)	(0.018)	(0.014)
Observations	491,871	448,537	448,537	448,537	448,537
R-Squared	0.201	0.203	0.221	0.226	0.247
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.6 Sensitivity of DM results to including a n-day post-earnings-announcement window. Estimates of β from:
$DM_{i,t}^{n} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
Where $DM_{i,t}^{n}$ is a measure of the pre-earnings drift that includes n days after the earnings announcement in the return attributed to the earnings day itself. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

Figure C.2 shows that the time-series trends in QVS are robust to choices of n up to 5. Table C.7 shows that including up to 5 days does not quantitatively or qualitatively change my baseline results.

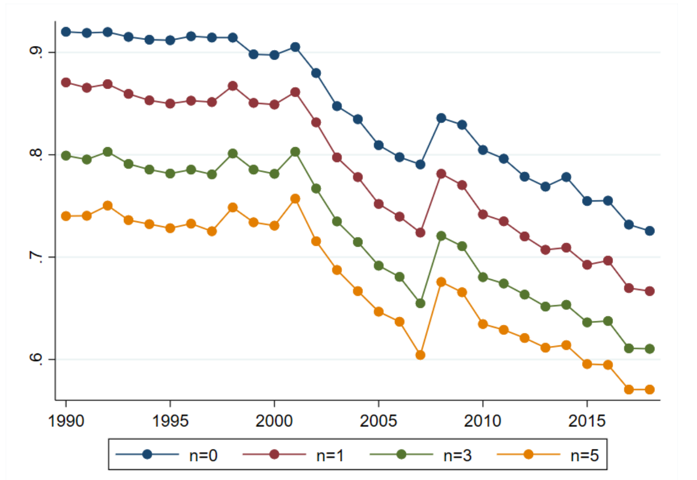

Figure C.2. Time series trends in QVS^n. This figure plots coefficients from a regression of QVS^n on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient.

C.6 Alternative explanations for the decline of price informativeness

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

C.6.1 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 leakage could increase the share of earnings information incorporated into prices before it was formally announced. 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.

	Include t+1				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.568***	-0.526***	-0.389***	-0.236**	-0.245***
	(0.026)	(0.028)	(0.030)	(0.101)	(0.088)
Observations	495,316	450,152	450,152	450,152	450,152
R-Squared	0.214	0.22	0.222	0.235	0.236
	Include t+1 to t+3				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.512***	-0.475***	-0.362***	-0.198**	-0.194**
	(0.026)	(0.028)	(0.029)	(0.096)	(0.085)
Observations	495,316	450,152	450,152	450,152	450,152
R-Squared	0.188	0.194	0.196	0.219	0.221
	Include t+1 to t+5				
	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.465***	-0.431***	-0.337***	-0.168*	-0.160*
	(0.024)	(0.026)	(0.027)	(0.094)	(0.083)
Observations	495,316	450,152	450,152	450,152	450,152
R-Squared	0.165	0.172	0.174	0.202	0.205
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.7 Sensitivity of QVS results to including a n-day post-earnings-announcement window. Table with estimates of β from:
$QVS^n_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where $QVS^n_{i,t}$ is a measure of earnings days' share of volatility which includes n days after the earnings announcement in the volatility attributed to the earnings announcement itself. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

Reg FD could be driving the trends in decreased price informativeness, as there was a large negative shock to information released by firms after it was passed. All of the information 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 the 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 before/after Reg FD was passed.

To further rule out this channel, I re-run the cross-sectional regressions using only post-2000 data in Tables C.8 and C.9. The results are qualitatively similar, which alleviates concerns that my results being driven by Reg FD.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.0353***	-0.0382***	-0.0432***	-0.0299	-0.0310**
	(0.007)	(0.007)	(0.007)	(0.019)	(0.015)
Observations	283,401	262,518	262,518	262,518	262,518
R-Squared	0.2	0.204	0.218	0.245	0.264
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.8 Passive Ownership and Pre-Earnings Drift (2001-2018). Table with estimates of β from:

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

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

Another possibility is that Reg FD changed the way insiders (e.g., directors or senior officers) behaved, or led to changes in the enforcement of insider trading laws (see e.g., Coffee (2007)). Time-series changes in enforcement should be accounted for by year fixed-

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.471***	-0.434***	-0.343***	-0.149	-0.236**
	(0.029)	(0.032)	(0.035)	(0.127)	(0.098)
Observations	284,156	262,999	262,999	262,999	262,999
R-Squared	0.205	0.207	0.209	0.184	0.185
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.9 Passive Ownership and Earnings Day Share of Volatility (2001-2018). Table with estimates of β from:

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

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

effects. To rule out the insider behavior channel, I used the Thompson Insiders data to compute insider buys/sells as a percent of total shares outstanding for each firm in my dataset. Insider buys and sells have been decreasing since the mid-1990's. Both average annual buys and sells went down slightly more for stocks with increases in passive ownership, but this effect is only weakly statistically significant. I then examined insider buys/sells in 22-day windows before/after earnings announcements. Both buys and sells have decreased before and after earnings announcements, broadly following the trend toward decreased insider activity. There is no statistically significant relationship, however, between passive ownership and insider buys/sells before or after earnings announcements.

C.6.2 The Rise of algorithmic trading (AT) activity

Weller (2018) shows that Algorithmic Trading (AT) activity is negatively correlated with pre-earnings price informativeness. The proposed mechanism is algorithmic traders back-run informed traders, reducing the returns to gathering firm-specific fundamental information.

AT activity increased significantly over my sample period, and could be responsible for some of the observed decrease in 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 cannot directly observe AT activity, and only have AT activity proxies between 2012-2018. I can, however, measure the effect of AT activity on the cross-sectional regression results. For AT activity to influence the regression estimates, it would have to be correlated with passive ownership, which I find plausible because: (1) Passive ownership is higher in large, liquid stocks, where most AT activity occurs. This, however, should not affect my results, as I condition on firm size in all the cross-sectional regressions and (2) High ETF ownership will attract algorithmic traders implementing ETF arbitrage. The effect of time trends in AT activity should be absorbed by the year fixed effects.

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

I re-run all the cross-sectional regressions, but restrict to the years with matched MIDAS data: 2012-2018. I then add the 4 AT activity measures to $X_{i,t}$ i.e., the controls to determine whether they reduce the ability of passive ownership to explain decreases in price informativeness. Tables C.10 and C.11 contain the results. The pre-earnings drift estimates are qualitatively and quantitatively similar when (1) restricting to the 2012-2018 sample i.e., columns 1, 2 and 4 and (2) including the AT controls i.e., columns 3 and 5. This is also true for the QVS regressions. These tables jointly suggest that the correlation between passive ownership and AT activity is not driving my results.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.0475***	-0.0462***	-0.0492***	-0.0448	-0.0601***
	(0.011)	(0.010)	(0.011)	(0.034)	(0.020)
Observations	94,529	80,989	80,989	80,989	80,989
R-Squared	0.251	0.261	0.267	0.295	0.306
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.10 Passive Ownership and Pre-Earnings Drift (Algorithmic Trading). Table with estimates of β from:

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

Where $DM_{i,t}$ is a measure of the pre-earnings drift. $Y_{i,t}$ contains the four measures of algorithmic trading (AT) activity from Weller (2018): (1) odd lot ratio, (2) trade-to-order ratio, (3) cancel-to-trade ratio and (4) average trade size. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

C.6.3 Effect of passive ownership on corporate governance

Given the literature on the effects of passive owners on corporate governance (see e.g., Appel et al. (2016)), one could worry that passive ownership's primary affect is to change governance, and then governance changes price informativeness. One mechanism would be that better governance leads to fewer information leaks, which in turn makes prices less informative before earnings announcements.[2]

To test this, I quantify corporate governance using the entrenchment index (E index) of Bebchuk et al. (2009). Using data from ISS between 1990 and 2018[3], I calculate this as the sum of indicator variables for the presence of: (1) a staggered (classified) board (2)

[2]There is, however, mixed evidence on this. For example, quoting Gloßner (2018), "I also find that passive investors have no significant effect on corporate social responsibility (CSR) ...", and the measure of CSR he uses includes corporate governance.

[3]Data from 1990-2006 is in a separate database – "ISS – Governance Legacy" – than the data from 2007 onward.

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	-0.380***	-0.309***	-0.318***	-0.264	-0.440***
	(0.044)	(0.043)	(0.050)	(0.188)	(0.145)
Observations	94,717	85,588	85,588	85,588	85,588
R-Squared	0.242	0.243	0.246	0.19	0.198
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table C.11 Passive Ownership and Earnings Day Share of Volatility (Algorithmic Trading). Table with estimates of β from:

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

where $QVS_{i,t}$ is a measure of earnings days' share of volatility. $Y_{i,t}$ contains the four measures of algorithmic trading (AT) activity from Weller (2018): (1) odd lot ratio, (2) trade-to-order ratio, (3) cancel-to-trade ratio and (4) average trade size. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

a limitation on amending bylaws (3) a limitation on amending the corporate charter (4) a requirement of a supermajority to approve a merger (5) golden parachutes for management/board members and (6) a poison pill. I then run a regression of the E index on passive ownership. Given that the E index is only defined annually, I use end of year data for passive ownership as well as all the control variables. Also, given that coverage is not equally spread across the firm size distribution, I do not report the value-weighted regression results.

Consistent with Gloßner (2018), table C.12 shows there isn't a statistically significant relationship between governance and passive ownership. The effect of a 15% increase in passive ownership on the E index is less than 0.05, so the effect of passive ownership on governance is also economically small relative both to the mean (≈3) and the standard deviation (≈1.5). I also find that my baseline regressions are unchanged by explicitly controlling for the E index. Jointly, this evidence suggests that the relationship between passive ownership and corporate governance is not driving my results.

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

Table C.12 Passive ownership and entrenchment. Table with estimates of β from:
$EIndex_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where $EIndex$ is the entrenchment index of Bebchuk et al. (2009). Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

D Robustness of quasi-experimental results

D.1 S&P Index Addition Details

The treatment effect of being added to the S&P 500 index has been increasing over time. Figure D.1 shows the average increase in passive ownership by year. In the early 1990s, firms experienced less than a 1% increase in passive ownership after being added to the index, while now the increase is over 3%.

D.2 Russell Details

D.2.1 May Market Capitalizations and Treated/Control Firms

I use the following procedure, based on Chang et al. (2015) and Coles et al. (2020), 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:

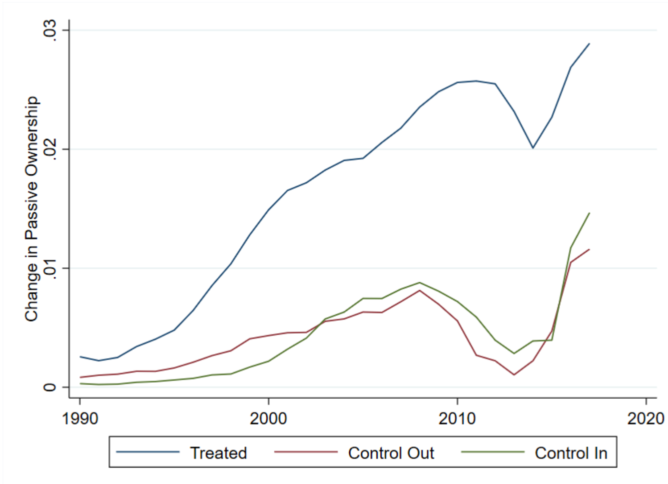

Figure D.1. Increase in Passive Ownership from S&P 500 Index Addition. 5-year moving average of increase in passive ownership associated with index addition. Increase in passive ownership is computed from 6 months before index addition to 6 months after index addition.

- 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.[4]
- 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 capi-

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

talization 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.

  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. (2020) 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.[5]
- 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

[5] 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. (2020) and they find using the total market capitalization of the 3000 vs. 3000E makes almost no difference to the accuracy of predicted index membership.

D.3 Increase in passive ownership when switching from the 1000 to the 2000

The treatment effect of switching from the Russell 1000 to 2000 has been increasing over time. Figure D.2 shows the average increase in passive ownership by year. In 2007, firms switching to the 2000 experienced a less than 2% increase in passive ownership, while now the increase is over 4%.

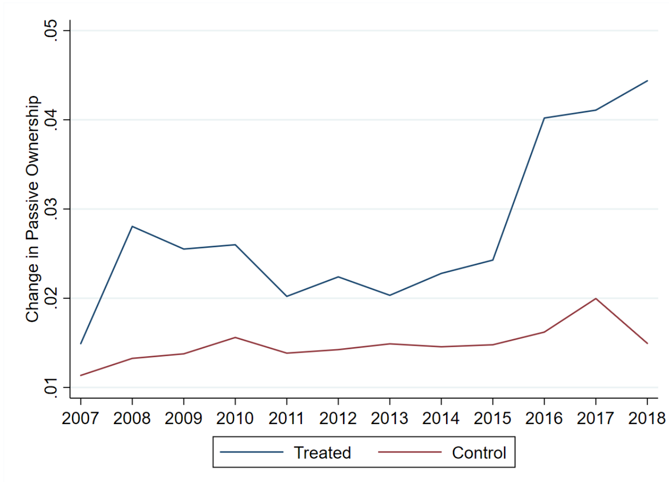

Figure D.2. Increase in Passive Ownership from Russell 1000/2000 Reconstitution. 5-year moving average of increase in passive ownership associated with switching from the Russell 1000 to the Russell 2000. Increase in passive ownership is computed from 6 months before index rebalancing to 6 months after index rebalancing.

D.4 Alternative quasi-exogenous changes in passive ownership

D.4.1 Moving from the Russell 2000 to the Russell 1000

Firms experience a decrease in passive ownership after they are moved from the Russell 2000 to the Russell 1000. This is because 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.

Again, following Coles et al. (2020), I choose the control firms to be those within $\pm$ 100 ranks of the upper band that were in the Russell 2000 the previous year. Figure D.3 shows the problem with this setup: the treatment is small and temporary. Within 12 months of switching, passive ownership is almost back at the pre index-rebalancing level.

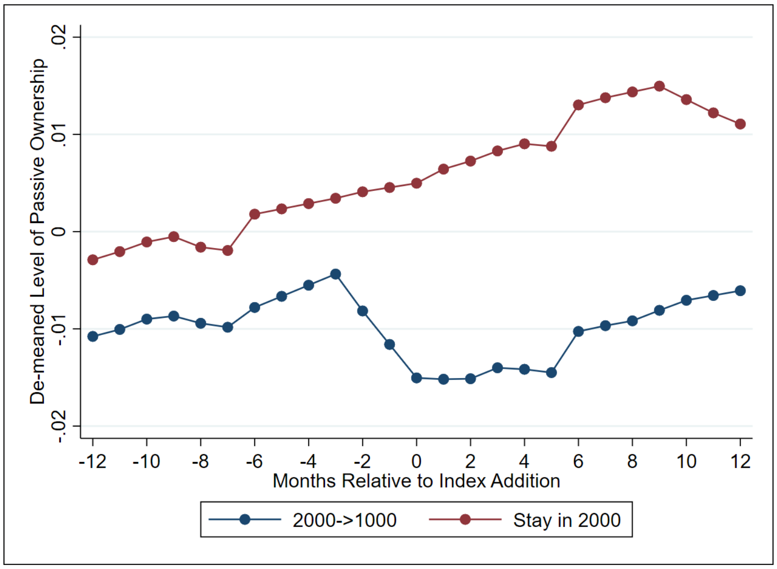

Figure D.3. Russell 1000/2000 Reconstitution and Changes in Passive Ownership. Average level of passive ownership for firms that stay in the Russell 2000 (control firms) and firms that moved from the Russell 2000 to the Russell 1000 (treated firms). Passive ownership is demeaned within each cohort.

D.4.2 Blackrock's acquisition of Barclays Global Investors

Another well-known source of quasi-exogenous variation in passive ownership is 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 (see e.g., Azar et al. (2018), Massa et al. (2021)) (2) While there may have been a *relative* increases in flows to iShares ETFs, relative to all other ETFs (see e.g., 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 investors, my proposed mechanism has no predictions for the effect of moving dollars from iShares ETFs to non-iShares ETFs.

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

In tables 4 and 5, the IV regressions are highly significant, while the reduced form regressions are sometimes insignificant. The concern is that, as discussed in Chernozhukov and Hansen (2008), an insignificant reduced form is evidence of weak instruments. In their notation:

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

Specifically, suppose the instruments are weak so $cov(Z, X)$ is close to zero. Then $(Z'Y) / (Z'X)$ i.e., the IV estimate of β might be large, but not because the true β is large. They argue that another way to test whether the true $\beta = 0$ is to check if $\gamma = 0$ i.e., the reduced form is insignificant.[6] At a high level, this is likely not a problem in my setting, as the first stage is very strong ($F > 300$ for the S&P experiment and $F > 100$ for the Russell experiment).

To better understand why the IV is significant, but the reduced form is insignificant, consider the case of a univariate structural regression and a single instrument. The model is

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

For simplicity, assume all variables are mean zero and that we have *iid* sampling so a standard LLN and CLT hold. Further, assume that $E[z_i\varepsilon_i] = 0$ so IV is consistent, 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: $E[x_i\varepsilon_i] = E[(\gamma z_i + u_i)\varepsilon_i] = E[u_i\varepsilon_i]$

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

OLS is inconsistent: We have

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

Because of the correlation between x_i and ε_i

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

IV is consistent: The IV estimator is

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

The exclusion restriction implies $E[\varepsilon_i z_i] = 0$ so we have $\widehat{\beta}_{IV} - \beta \xrightarrow{p} 0$ i.e., the IV estimator is consistent and

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

Reduced form is consistent: We have

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

Thus, we have $\widehat{\alpha}_{RF} - \alpha \xrightarrow{p} 0$ i.e., the reduced form is consistent and

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

Distribution of centered t-statistics: Assuming homoskedasticity, we have

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

And

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

We can also write these as

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

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

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

If the true $\beta = 0$, we have $\rho_{\varepsilon,v} = 1$ and the t-statistics will be perfectly correlated asymptotically. Crucially, if $\beta \neq 0$, then we have $\rho_{\varepsilon,v} < 1$ and these two t-statistics will not be perfectly correlated, even asymptotcically. Thus, it is possible to have a significant IV estimate and insignificant reduced form estimate. This is more like when $\rho_{\varepsilon,v}$ is smaller.

Distribution of uncentered t-statistics: If we compute these t-statistics under the $\alpha = \beta = 0$ null as in Lochner and Moretti (2004), we have

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

and taking their ratio yields:

$$\frac{t_{\widehat{\beta}_{IV}}}{t_{\widehat{\alpha}_{RF}}} = \frac{\sqrt{(N^{-1}\sum_i \widehat{v}_i^2)}}{\sqrt{(N^{-1}\sum_i \widehat{\varepsilon}_i^2)}} \xrightarrow{p} \sqrt{\frac{E\left[v_i^2\right]}{E\left[\varepsilon_i^2\right]}} = \sqrt{\frac{\beta^2 E\left[u_i^2\right] + 2\beta E\left[u_i\varepsilon_i\right] + E\left[\varepsilon_i^2\right]}{E\left[\varepsilon_i^2\right]}} \tag{4}$$

Thus, when we compute the t-statistics under the $\beta = \alpha = 0$ null, they will be perfectly correlated asymptotically. Of course, 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. However, these two t-statistics will still be perfectly correlated in large samples.

In reality, we do not know α and β, so we cannot compute the centered t-statistics. Empirically, in tables 4 and 5, I am testing whether $\widehat{\beta}_{IV} = 0$ and $\widehat{\alpha}_{RF} = 0$. With this in mind, assuming the true β and α are not zero, there are two things to consider:

1. In my setting, I expect $\beta < 0$ i.e., passive ownership decreases price informativeness. If this is the case, as the covariance between u_i and ε_i becomes more negative, we expect the uncentered t-statistic for the IV to be relatively larger than the uncentered t-statistic for the reduced form. This is because this increasing negative covariance will tend to increase $\beta E\left[u_i\varepsilon_i\right]$ in the numerator of Equation 4, increasing the ratio of the IV t-statistic to the reduced form t-statistic.
2. 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
   $\hat{\sigma}_v^2 = (y_i - \hat{\alpha}_{RF}z_i)'(y_i - \hat{\alpha}_{RF}z_i)/N$

Simulation results: I simulate the above model, varying the sign and the strength of $E\left[u_i\varepsilon_i\right]$. Given that the sample size matters, I choose $N = 10,000$ to match the number of observations in table 5. I set $\beta = -0.25$, $\gamma = 0.5$, although all results are similar using any $\beta < 0$ and $\gamma \neq 0$. u is independent from everything else, and distributed standard normal. I consider 3 versions of ε: (1) $\varepsilon_1 = n$ where n is an *iid* standard normal random variable (2) $\varepsilon_2 = -1 \cdot \phi \cdot u + n$ where $\phi > 0$ is a constant which controls the covariance between ε_2 and u (3) $\varepsilon_3 = \phi \cdot u + n$, which is similar to ε_2, except now u and ε_3 will be positively correlated instead of negatively correlated. Finally, to ensure that everything is not statistically significant in every simulation, I add additional noise, multiplying every ε by 3 and u by 5.

Figure D.4 plots the fraction of simulations where the t-statistic from the IV is less than -1.96, but the t-statistic from the RF is greater than -1.96. The blue dots show that in case 1, the RF is slightly less likely to be statistically significant, which is not surprising as even if $E[u_i \varepsilon_i] = 0$, the ratio in Equation 4 will be bigger than one. In case 2 we see that the RF is even less significant on average than the IV, and this effect is increasing in strength of the negative correlation between u and ε. Finally, in case 3, we see the opposite. As the covariance between u and ε increases, the numerator shrinks, as this positive covariance is being multiplied by β, which is less than 0.

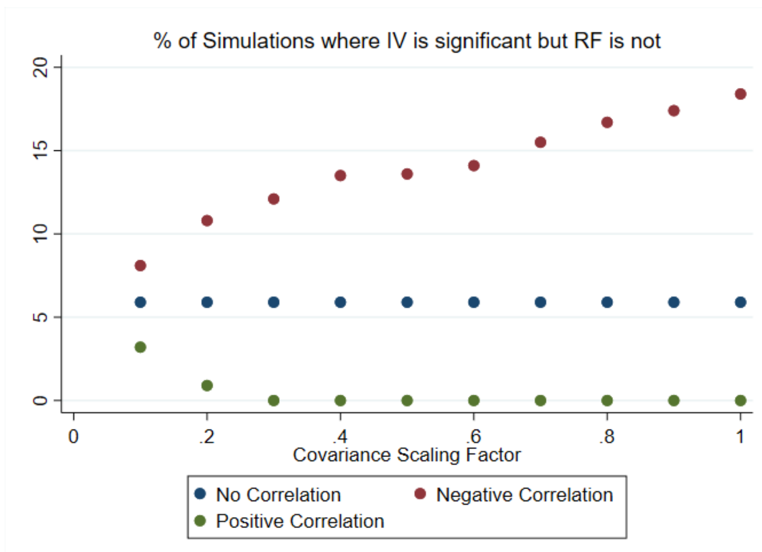

Figure D.4. Comparison of statistical significance. Each dot represents the percentage of simulations where the instrumental variables specification is statistically significant, but the reduced form is not. The blue dots are from simulations where ε is independent of u, the red dots are from simulations where ε is negatively correlated with u and the green dots are from simulations where ε is positively correlated with u. Moving from left to right increases the (absolute) covariance between ε and u for the red and green dots.

Economic mechanism: The natural next question is whether there is an economic reason why the reduced form is insignificant but the IV is significant. As shown in Figure D.4, this is more likely to happen if the true β is negative and the covariance between u_i

and ε_i is sufficiently negative.

In terms of the economic meaning of $E[u_i \varepsilon_i] < 0$, consider the following: 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 closet indexers, $closet_{i,t}$ (i.e., funds which are passive, but don't explicitly say so). Suppose further that the data generating process for price informativeness is:

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

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 closet indexing is proportional to observed indexing i.e., $closet_{i,t} = \psi \cdot passive_{i,t}$ where $\psi > 0$. This might be the case if e.g., there are closet indexers who also track the sub-indices.

Now, in the continuous instrument case, 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 $PassiveGap_{i,t}$. Recall the first stage regression:

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

In this case, $u_{i,t}$ would be positive, because firm i got a larger than expected expected increase in passive ownership because they were 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

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

where $\tilde{\varepsilon}_{i,t} = \varepsilon_{i,t} + \beta \cdot closet_{i,t}$.

Now, in this setting $u_{i,t}$ and $\tilde{\varepsilon}_{i,t}$ are going to have negative covariance, because $closet_{i,t}$ is positively related to $passive_{i,t}$. And if $\beta < 0$, then $\beta E\left[u_i \varepsilon_i\right] > 0$, which would tend to make the reduced form have a smaller t-statistic than to the IV.

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

D.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. 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 total institutional ownership.

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

Gloßner's results for Russell reconstitutions end in 2006, and I am using the Russell experiment from 2007 onward, so to make sure his conclusion also applies in my setting, I expand his results to 2020. I also want to check for changes in total institutional ownership around S&P 500 index additions. To this end, each month, I compute the value-weighted average level of total institutional ownership and passive ownership. Then, I identify months $t = -6$ to $t = 6$ around each index reconstitution event (these results, however, are not sensitive to this choice of a $\pm$ 6-month window). Finally, I calculate the total change in passive and institutional ownership between $t = -6$ and $t = 6$ and subtract the change in

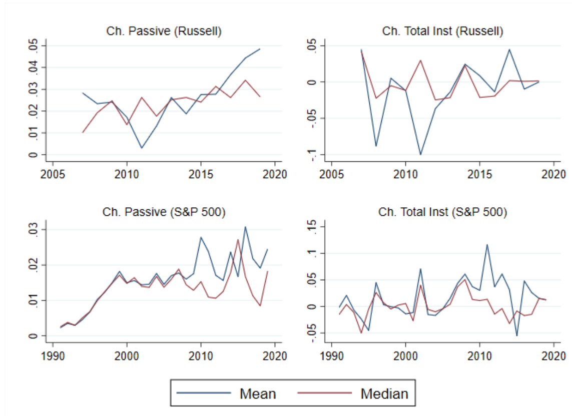

Figure D.5. Changes in ownership around index reconstitution events. For each firm which is added to the S&P 500 or that switches from the Russell 1000 to the Russell 2000, I compute the change in passive ownership and institutional ownership between month $t-6$ and month $t+6$. Then, I subtract the change in the corresponding value-weighted average ownership over the same period. This figure plots the de-trended mean and median change in each ownership type for the treated firms.

the corresponding value-weighted average over the same period to remove aggregate trends.

Figure D.5 contains the results. When a firm switches from the Russell 1000 to the 2000 (top two panels), on average, they have experienced an increase in passive ownership every year in my sample. These same firms, however, do not have a consistent increase in institutional ownership. Some years institutional ownership increases, while in other years it decreases. Similarly, when a firm is added to the S&P 500 (bottom two panels), they receive an increase in passive ownership every year in my sample. As with the Russell reconstitution, however, the change in institutional ownership is sometimes positive and sometimes negative.

While the results in Figure D.5 initially seem counter intuitive, they make sense when we think about how we expect retail (i.e., everyone who is not an institutional investor) to trade around index events. For example, when FootLocker was dropped from the S&P 500 in 8/2019, I find it unlikely that retail investors (1) read the press release on S&P's website and (2) even if they did read the press release, decided to sell their shares to institutions.

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 regressions, I can and do explicitly control for total institutional ownership. In fact, I find there is significant cross-sectional variation of passive ownership within various levels of institutional ownership. For example, Figure D.6 plots passive ownership against institutional ownership in 12/2018. While these two quantities are correlated (R-squared of about 50%) – and they should be, because passive ownership is included in total institutional ownership – they are far from perfectly correlated.

In addition, in table D.1, I replicate all the instrumental variables and reduced form regressions, including total institutional ownership on the right hand side. All the results are quantitatively unchanged from tables 4 and 5 in the main body of the paper.

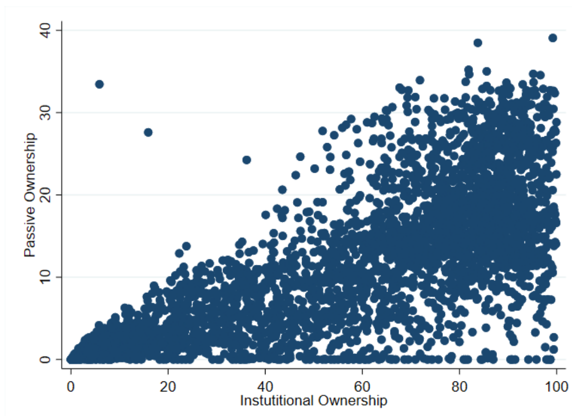

Figure D.6. Passive Ownership vs. Institutional Ownership. Scatterplot of passive ownership against total institutional ownership. Each blue dot represents a single firm-quarter observation.

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

	Panel A: Binary Instrument (Russell)							Panel C: Binary Instrument (S&P)						
		Pre-Earnings Volume		Pre-Earnings Drift		QVS			Pre-Earnings Volume		Pre-Earnings Drift		QVS	
	First Stage	IV	RF	IV	RF	IV	RF	First Stage	IV	RF	IV	RF	IV	RF
Post x Treated	0.0117***		-0.515		-0.00627**		-0.00959	0.0134***		-0.563***		-0.00127		-0.00514
	(0.004)		(0.448)		(0.003)		(0.012)	(0.001)		(0.199)		(0.001)		(0.005)
Passive Ownership		-22.12***		-0.100***		-0.398***			-24.63***		-0.138***		-1.958***	
		(6.818)		(0.032)		(0.122)			(8.391)		(0.039)		(0.149)	
Observations	9,332	9,332	9,332	9,794	9,794	9,805	9,805	279,604	279,604	279,604	283,986	283,986	285,367	285,367
F-statistic	165							304						
	Panel B: Continious Instrument (Russell)							Panel D: Continious Instrument (S&P)						
		Pre-Earnings Volume		Pre-Earnings Drift		QVS			Pre-Earnings Volume		Pre-Earnings Drift		QVS	
	First Stage	IV	RF	IV	RF	IV	RF	First Stage	IV	RF	IV	RF	IV	RF
Post x Treated	0.865***		-5.917		-0.224		-0.433	0.544***		-22.83***		-0.00719		-0.177
x Passive Gap	(0.200)		(18.160)		(0.138)		(0.582)	(0.045)		(7.792)		(0.035)		(0.162)
Passive Ownership		-21.61***		-0.101***		-0.398***			-24.67***		-0.135***		-1.951***	
		(6.595)		(0.031)		(0.122)			(8.360)		(0.039)		(0.150)	
Observations	9,332	9,332	9,332	9,794	9,794	9,805	9,805	279,604	279,604	279,604	283,986	283,986	285,367	285,367
F-statistic	183							420						
Cross-sectional regression estimate		-11.49***	-11.49***	-0.0480***	-0.0480***	-0.395***	-0.395***		-11.49***	-11.49***	-0.0480***	-0.0480***	-0.395***	-0.395***

Table D.1Quasi-experimental results including total institutional ownership on the right-hand-side

Estimates from:

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

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

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

Column 1 in each panel is a first-stage regression. Columns 2 and 4 are instrumental variables regressions. Columns 3 and 5 are reduced-form regressions. Panels A and C contain regressions from the binary treatment specification, while Panels B and D contain regressions from the continuous treatment specification. Panels A and B correspond to the Russell experiment, while Panels C and D correspond to the S&P experiment. FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

E Mechanisms Details

E.1 Effects of ETFs vs. passive index funds on price informativeness

As discussed in the paper, if pre-earnings announcement prices have become less informative, the returns to becoming informed should have increased. So, in general equilibrium, it's not obvious why the remaining non-passive investors wouldn't increase their information production to offset this.

A natural reason why non-passive investors wouldn't fully compensate for the decline in information production is that passive ownership's presence makes it harder to profit from private information. As discussed in Ben-David et al. (2018), ETFs – but not non-ETF index funds – increase non-fundamental volatility in the underlying stocks. This could deter informed investors from gathering information, as there is some chance that before the end of their investment horizon, they are hit with a large volatility shock, which forces them to sell at a loss.

An implication of this is that the effects of passive ownership on price informativeness should be stronger for ETFs than non-ETF passive funds. To test this, I re-run my baseline regressions, but break institutional ownership into four mutually exclusive pieces: (1) ETFs (2) All passive funds that aren't ETFs (3) All mutual funds i.e., S12 filers which aren't passive and (4) All institutions i.e., 13F filers which are not mutual funds. Consistent with this, table E.1 shows that ETFs have a larger effect on $DM_{i,t}$ and $QVS_{i,t}$ than passive mutual funds.

One concern with the results in table E.1 is that ETFs explain all the effects of passive ownership on price informativeness. Separating the effects of index mutual funds from ETFs is difficult, however, as they have a correlation coefficient of almost 0.7. To check this, in table E.2, I replicate table E.1, but exclude ETF ownership from the right-hand-side. This restores the statistical significance of index/passive mutual fund ownership, confirming that my results are not entirely driven by ETFs.

	LHS: DM				
	(1)	(2)	(3)	(4)	(5)
ETF	-0.0255***	-0.0313***	-0.0566***	-0.0898***	-0.0565***
	(0.010)	(0.010)	(0.008)	(0.023)	(0.019)
Non-ETF Index	-0.0144	-0.0282*	-0.0142	0.035	0.0161
	(0.015)	(0.015)	(0.013)	(0.031)	(0.027)
Active MF	0.0131***	0.0104***	-0.00429*	0.0126**	0.00861*
	(0.003)	(0.003)	(0.002)	(0.005)	(0.005)
Other Inst.	0.00903***	0.00897***	0.00406***	0.00525**	0.00397
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Observations	345,831	333,013	333,013	333,013	333,013
R-Squared	0.207	0.209	0.225	0.245	0.263
	LHS: QVS				
	(6)	(7)	(8)	(9)	(10)
ETF	-0.671***	-0.662***	-0.554***	-0.297**	-0.285*
	(0.050)	(0.051)	(0.052)	(0.142)	(0.148)
Non-ETF Index	-0.04	-0.0413	-0.0139	0.078	0.0161
	(0.063)	(0.065)	(0.065)	(0.208)	(0.198)
Active MF	-0.0817***	-0.0794***	-0.0352***	0.0571*	0.0305
	(0.011)	(0.012)	(0.013)	(0.034)	(0.038)
Other Inst.	0.00377	0.00469	0.0123**	0.012	0.00263
	(0.005)	(0.006)	(0.006)	(0.015)	(0.015)
Observations	345,537	332,708	332,708	332,708	332,708
R-Squared	0.221	0.222	0.223	0.23	0.231
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table E.1 Breakdown of DM and QVS regression results by ownership type. Table with estimates of b_is from:

$PriceInformativeness_{i,t} = \alpha + b_1 ETF_{i,t} + b_2 NonETFIndex_{i,t} + b_3 ActiveMF_{i,t} + b_4 OtherInst_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$

where $PriceInformativeness_{i,t}$ is either DM or QVS. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All ownership measures are expressed as decimals, so 0.01 = 1% of firm i's shares are owned by ETFs. Standard errors double clustered at the firm and year-quarter level in parenthesis.

	LHS: DM				
	(1)	(2)	(3)	(4)	(5)
Non-ETF Index	-0.0309**	-0.0476***	-0.0488***	-0.0464	-0.0345
	(0.015)	(0.014)	(0.013)	(0.036)	(0.032)
Active MF	0.0136***	0.0110***	-0.00258	0.0150***	0.0102**
	(0.003)	(0.003)	(0.003)	(0.005)	(0.005)
Other Inst.	0.00966***	0.00974***	0.00550***	0.00684***	0.00491**
	(0.002)	(0.002)	(0.001)	(0.002)	(0.002)
Observations	346,089	333,267	345,831	333,013	333,013
R-Squared	0.047	0.047	0.207	0.209	0.225
	LHS: QVS				
	(6)	(7)	(8)	(9)	(10)
Non-ETF Index	-0.472***	-0.451***	-0.266***	-0.191	-0.181
	(0.069)	(0.070)	(0.071)	(0.230)	(0.202)
Active MF	-0.0685***	-0.0662***	0.0154	0.0650**	0.0517
	(0.012)	(0.012)	(0.013)	(0.032)	(0.035)
Other Inst.	0.0203***	0.0210***	0.0330***	0.0172	0.00972
	(0.005)	(0.005)	(0.006)	(0.015)	(0.015)
Observations	345,796	332,963	345,537	332,708	332,708
R-Squared	0.114	0.115	0.219	0.22	0.222
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table E.2 Breakdown of DM and QVS regression results by ownership type, excluding ETFs. Table with estimates of b_is from:
$PriceInformativeness_{i,t} = \alpha + b_1 NonETFIndex_{i,t} + b_2 ActiveMF_{i,t} + b_3 OtherInst_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where $PriceInformativeness_{i,t}$ is either DM or QVS. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All ownership measures are expressed as decimals, so 0.01 = 1% of firm i's shares are owned by ETFs. Standard errors double clustered at the firm and year-quarter level in parenthesis.

E.2 Earnings Responses

I aim to quantify market responses to earnings information of a given size. To measure trends in earnings responses, I run two types of regressions. The baseline comes from Kothari and Sloan (1992):

$$r_{i,t} = \alpha + \beta SUE_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \tag{5}$$

where $r_{i,t}$ denotes the market-adjusted return on the effective quarterly earnings date i.e., the first day investors could trade on earnings information. $r_{i,t}$ is Winsorized at the 1% and 99% level by year. $SUE_{i,t} = \frac{E_{i,t}-E_{i,t-4}}{\sigma_{(t-1,t-8)}(E_{i,t}-E_{i,t-4})}$ where $E_{i,t}$ is earnings-per-share from the IBES unadjusted detail file i.e., "street" earnings, so the numerator is the year-over-year (YOY) earnings growth, while the denominator is the standard deviation of YOY earnings growth over the past 8 quarters. I compute SUE this way, following Novy-Marx (2015), because it avoids (1) using prices as an input, whose average informativeness has changed over time and (2) using analyst estimates of earnings as an input, whose average accuracy has also changed over time. As a result, the average absolute value of $SUE_{i,t}$ is roughly constant over my sample, except for large spikes during the tech boom/bust as well as during the global financial crisis.

I also design an earnings-response regression which allows for asymmetry between positive and negative surprises:

$$r_{i,t} = \alpha + \beta_p 1_{SUE_{i,t} \geq 0} \times SUE_{i,t} + \beta_n 1_{SUE_{i,t} < 0} \times |SUE_{i,t}| + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \tag{6}$$

I run these regressions in 5-year rolling windows and plot the β's in Figure E.1. Over the past 30 years, earnings responses have increased by a factor of over 3×. Most of this increase was driven by increased responsiveness to SUEs greater than zero. In recent years, however, this trend has reversed, with the response to positive news decreasing and the response to negative news increasing.

One concern with the earnings-response results is that they are specific to only including the return on the effective earnings announcement date itself on the left-hand-side. To

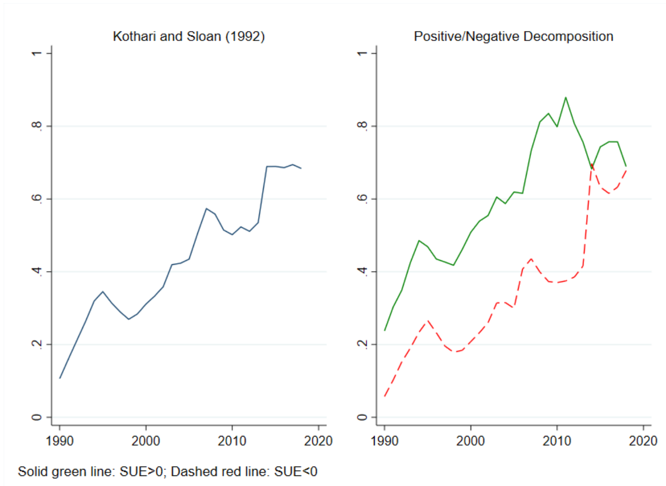

Figure E.1. Trends in Earnings Response. Left panel has estimates of β from:

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

run in 5-year rolling windows. Right panel has estimates of β_1 and β_2 from Equation 6 i.e., breaking SUE into positive and negative components. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All regressions contain year-quarter fixed effects, ϕ_t and firm fixed effects ψ_i.

alleviate this concern, I run the following regression

$$\begin{aligned} r_{i,(t,t+n)} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 \left(SUE_{i,t} \times Passive_{i,t}\right) \\ + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \end{aligned} \tag{7}$$

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

Table E.3 shows that even including up to 5 days in $r_{i,(t,t+n)}$ does not change that (1) high passive stocks respond more to earnings information than low passive stocks (2) this effect is stronger for negative news and (3) this effect is even stronger for negative idiosyncratic news.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
		$n=1$			$n=3$			$n=5$	
SUE	0.358***			0.472***			0.493***		
	(0.019)			(0.024)			(0.026)		
$SUE \times 1_{SUE>0}$		0.583***			0.790***			0.848***	
		(0.030)			(0.036)			(0.040)	
$\lvert SUE\rvert \times 1_{SUE\leq 0}$		-0.227***			-0.287***			-0.285***	
		(0.019)			(0.024)			(0.026)	
$Sys.SUE \times 1_{Sys.SUE>0}$			0.270**			0.254			0.253
			(0.133)			(0.194)			(0.251)
$\lvert Sys.SUE\rvert \times 1_{Sys.SUE\leq 0}$			-0.353**			-0.379*			-0.362
			(0.164)			(0.226)			(0.279)
$Idio.SUE \times 1_{Idio.SUE>0}$			0.593***			0.804***			0.866***
			(0.032)			(0.039)			(0.044)
$\lvert Idio.SUE\rvert \times 1_{Idio.SUE\leq 0}$			-0.245***			-0.309***			-0.304***
			(0.021)			(0.027)			(0.029)
SUE	2.874***			2.874***			2.960***		
$\times Passive$	(0.238)			(0.315)			(0.356)		
$SUE \times 1_{SUE>0}$		1.293***			0.693			0.449	
$\times Passive$		(0.373)			(0.470)			(0.509)	
$\lvert SUE\rvert \times 1_{SUE\leq 0}$		-3.896***			-4.260***			-4.590***	
$\times Passive$		(0.330)			(0.450)			(0.504)	
$Sys.SUE \times 1_{Sys.SUE>0}$			1.385			3.185**			3.852**
$\times Passive$			(1.006)			(1.421)			(1.770)
$\lvert Sys.SUE\rvert \times 1_{Sys.SUE\leq 0}$			0.667			1.858			2.399
$\times Passive$			(2.869)			(3.781)			(4.683)
$Idio.SUE \times 1_{Idio.SUE>0}$			1.038**			0.411			0.115
$\times Passive$			(0.433)			(0.545)			(0.588)
$\lvert Idio.SUE\rvert \times 1_{Idio.SUE\leq 0}$			-3.889***			-4.216***			-4.609***
$\times Passive$			(0.324)			(0.410)			(0.426)
Observations	354,861	354,861	333,875	354,861	354,861	333,875	354,861	354,861	333,875
R-squared	0.064	0.065	0.064	0.067	0.068	0.067	0.067	0.068	0.067

Table E.3 Sensitivity of earnings-response regressions to including a n-day post-earnings-announcement window. Estimates from:
$r_{i,(t,t+n)} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \gamma_1 \left(SUE_{i,t} \times Passive_{i,t}\right) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}$
where $r_{i,(t,t+n)}$ is the cumulative log market-adjusted return (in percentage points) from the effective earnings announcement date to $t+n$. $r_{i,(t,t+n)}$ is Winsorized at the 1% and 99% level by year. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All columns contain year-quarter fixed effects, ϕ_t, and firm fixed effects ψ_i. Standard errors double clustered at the firm and year-quarter level in parenthesis.

E.3 Option Implied Volatility

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 (see e.g., Beber and Brandt (2006)), b is selected so that it is at least 5 days after τ.[7]

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, compute $IV_{i,t,e}$, defined as the equal-weighted average implied volatility across all at-the-money options expiring on date e. Then, take an equal-weighted average of $IV_{i,t,b}$ over the 20-day window before τ:

$$\overline{IV}_{i,b} = Mean\left[IV_{i,(b-s,b),b} : b - s \in [\tau - 20, \tau - 1]\right] \tag{8}$$

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

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right) \tag{9}$$

higher values of $IVD_{i,\tau}$ imply that options which span earnings announcements are relatively more expensive i.e., there is more ex-ante uncertainty.[8]

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-

[7] 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.

[8] 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 $I\tilde{V}D_{i,\tau} = \overline{IV}_{i,b}/\frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right)$.

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.[9]

Figure E.2 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, on both an equal-weighted and value-weighted basis, IVD has increased by about 0.05 over the past 25 years. This is evidence that there is more uncertainty about fundamentals before earnings announcements now than there was in the late 1990's.

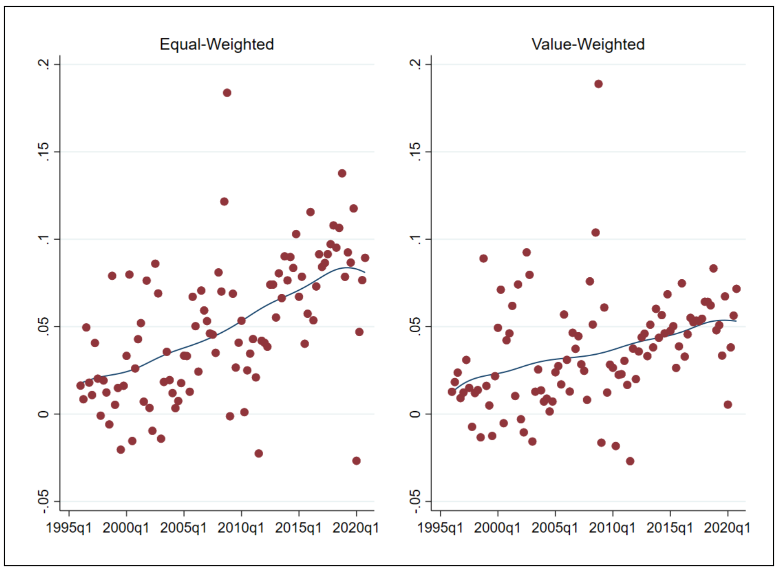

Figure E.2. Time-series trends in IVD. Equal-weighted and value-weighted averages of IVD by quarter. Red dots represent cross-sectional averages and blue lines represent LOWESS filters with bandwidths equal to 20% of quarters in the dataset.

E.4 Passive ownership and short interest

As discussed in section C.4 above, one reason for the asymmetric effect of passive ownership on the pre-earnings drift between firms with good and bad news could be that passive

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

	(1)	(2)	(3)	(4)	(5)
Passive Ownership	0.218***	0.231***	0.0876***	0.190***	0.0886***
	(0.013)	(0.017)	(0.014)	(0.020)	(0.017)
Observations	1,499,717	963,151	963,151	963,151	963,151
R-Squared	0.52	0.542	0.58	0.561	0.582
Firm + Year/Quarter FE	✓	✓	✓	✓	✓
Matched to Controls		✓	✓	✓	✓
Firm-Level Controls			✓		✓
Weight	Equal	Equal	Equal	Value	Value

Table E.4 Passive ownership and short interest. Estimates of β from:
$\overline{SR}_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where $\overline{SR}_{i,t}$ is the average short interest for stock i in month t. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All columns contain year-quarter fixed effects, ϕ_t, and firm fixed effects ψ_i. Standard errors double clustered at the firm and year-quarter level in parenthesis.

ownership affects short selling e.g., maybe passive owners are more willing to lend out their shares than other institutional investors. To test this, following Hanson and Sunderam (2014), I define the short interest ratio as $SR_{i,t} = SHORT_{i,t}/SHROUT_{i,t}$, where I obtain data on the number of shares shorted from Compustat and the number of shares outstanding from CRSP. This is reported every two weeks, but given that most of my variables are only defined monthly, I take the average $SR_{i,t}$ each month, which I call $\overline{SR}_{i,t}$. I then run regressions with $\overline{SR}_{i,t}$ on the left hand side and all the same variables as the baseline regressions on the right hand side. Table E.4 shows that stocks with more passive ownership tend to have more short interest on average. This continues to hold when controlling for other factors known to be correlated with ease of shorting (e.g., firm size and total institutional ownership, as discussed in Daniel et al. (2017)) and is true both on an equal-weighted and value-weighted basis.

As discussed in the paper, high passive ownership stocks have a bigger increase in short interest after they end up releasing bad news. I argue this is evidence that investors were relatively less informed about high passive stocks before the earnings announcement. An alternative explanation is that passive ownership has made it easier to short, so earnings information is incorporated into prices faster after it is released. If this were the case, however, we would expect that including a sufficiently long post-earnings announcement window would remove the drift asymmetry between firms that end up releasing good news and bad news.

To test this, I define an alternative measure of the pre-earnings drift, DM_{it}^n which includes n days after the effective announcement date in the earnings-announcement return:

$$DM_{it}^n = \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t+n)}} & \text{if } r_{(t,t+n)} > 0 \\ \frac{1+r_{(t-22,t+n)}}{1+r_{(t-22,t-1)}} & \text{if } r_{(t,t+n)} < 0 \end{cases} \tag{10}$$

If this alternative story is true, increasing n should decrease the asymmetry. One issue with using DM_{it}^n, however, is that the larger n is, the more other (non-earnings) information could be released. If other information comes out in the $t+1$ to $t+n$ window, then DM_{it}^n would not necessarily be capturing only the amount of earnings news in pre-announcement prices.

Table E.5 contains the results of re-running the pre-earnings drift regressions from table C.5, but putting DM_{it}^n on the left hand side for $n = 1$, $n = 3$ and $n = 5$ (an alternative way to think about table C.5, is that the left-hand-side is actually DM_{it}^n, with $n = 0$). The last row of the table presents the difference between the coefficient on $Passive \times LowSUE$ and $Passive \times HighSUE$. Inconsistent with the increased ease of shorting explanation, this difference remains large, even at $n = 5$.

E.5 Trends in pre-earnings turnover

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

$$AT_{i,t+\tau} = \alpha + \sum_{\tau=-21}^{22} \beta_\tau \mathbf{1}_{\{i,t+\tau\}} + e_{i,t+\tau} \tag{11}$$

The right-hand side variables of interest are a set of indicators for days relative to the earnings announcement. For example, $\mathbf{1}_{\{i,t-15\}}$ is equal to one 15 trading days before the nearest earnings announcement for stock i and zero otherwise. The regression includes all stocks in my sample and a ± 22 day window around each earnings announcement. Abnormal turnover is Winsorized at the 1% and 99% levels by year.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	n=1				n=3				n=5			
Passive	0.0366***	0.0327***	0.0032	-0.00407	0.0571***	0.0424***	0.000821	-0.0159	0.0653***	0.0454***	0.00851	-0.0124
	(0.009)	(0.009)	(0.017)	(0.014)	(0.010)	(0.009)	(0.017)	(0.013)	(0.010)	(0.008)	(0.019)	(0.017)
Low SUE	-0.0141***	-0.0123***	-0.0100***	-0.00893***	-0.0151***	-0.0128***	-0.0108***	-0.00964***	-0.0161***	-0.0132***	-0.0114***	-0.00974***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
2	-0.00356***	-0.00344***	-0.00268***	-0.00274***	-0.00402***	-0.00386***	-0.00227**	-0.00242**	-0.00419***	-0.00390***	-0.00238**	-0.00240**
	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
3	-0.00138***	-0.00151***	-0.000501	-0.000474	-0.00108***	-0.00113***	-0.000846	-0.000796	-0.00115***	-0.00115***	-0.000875	-0.000727
	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)
High SUE	-0.00855***	-0.00723***	-0.00506***	-0.00429***	-0.00862***	-0.00675***	-0.00484***	-0.00396***	-0.00891***	-0.00650***	-0.00456***	-0.00339***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Low SUE x Passive	-0.135***	-0.104***	-0.149***	-0.110***	-0.146***	-0.112***	-0.150***	-0.107***	-0.151***	-0.115***	-0.144***	-0.101***
	(0.014)	(0.012)	(0.023)	(0.018)	(0.016)	(0.013)	(0.026)	(0.021)	(0.017)	(0.014)	(0.027)	(0.023)
2 x Passive	-0.0432***	-0.0319***	-0.0455***	-0.0360***	-0.0446***	-0.0318***	-0.0497***	-0.0378***	-0.0432***	-0.0302***	-0.0466***	-0.0348***
	(0.006)	(0.005)	(0.009)	(0.009)	(0.006)	(0.005)	(0.011)	(0.010)	(0.006)	(0.005)	(0.011)	(0.011)
3 x Passive	-0.0362***	-0.0232***	-0.0272***	-0.0167*	-0.0398***	-0.0257***	-0.0206*	-0.0101	-0.0392***	-0.0237***	-0.0199	-0.00885
	(0.005)	(0.005)	(0.010)	(0.009)	(0.005)	(0.005)	(0.011)	(0.009)	(0.005)	(0.005)	(0.012)	(0.011)
High SUE x Passive	-0.103***	-0.0744***	-0.0600***	-0.0262**	-0.118***	-0.0896***	-0.0665***	-0.0305**	-0.125***	-0.0945***	-0.0740***	-0.0364***
	(0.010)	(0.008)	(0.014)	(0.012)	(0.011)	(0.009)	(0.015)	(0.013)	(0.012)	(0.010)	(0.016)	(0.013)
Observations	371,018	352637	371018	352637	370,994	352624	370994	352624	370,971	352609	370971	352609
R-squared	0.22	0.235	0.254	0.269	0.221	0.236	0.24	0.256	0.223	0.238	0.236	0.253
Firm + Year/Quarter FE	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Matched to Controls		✓		✓		✓		✓		✓		✓
Firm-Level Controls		✓		✓		✓		✓		✓		✓
Weight	Equal	Equal	Value	Value	Equal	Equal	Value	Value	Equal	Equal	Value	Value
High vs. Low x Passive	0.0322	0.0293	0.0895	0.0836	0.0279	0.0227	0.0834	0.0767	0.0255	0.0200	0.0702	0.0643

Table E.5 Passive ownership and pre-earnings drift, by SUE quintile, varying length of the earnings-return window . Estimates of β, b_j and c_j from:

$$DM^n_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^{5} b_j 1_{SUE_{i,t} \in Q_j} + \sum_{j=1}^{5} c_j 1_{SUE_{i,t} \in Q_j} \times Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$$

Where $DM_{i,t}$ is a measure of the pre-earnings drift that includes n days after the earning announcement in the return attributed to the earnings announcement itself. $1_{SUE_{i,t} \in Q_j}$ is an indicator for firm i being in the jth quintile of SUE at time t, where the quintiles are formed each quarter. For every regression, the middle quintile is the omitted group. Controls in $X_{i,t}$ include firm age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis. The last row is the p-value from a t-test that the coefficient on Low SUE x Passive is equal to the coefficient on High SUE x Passive.

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

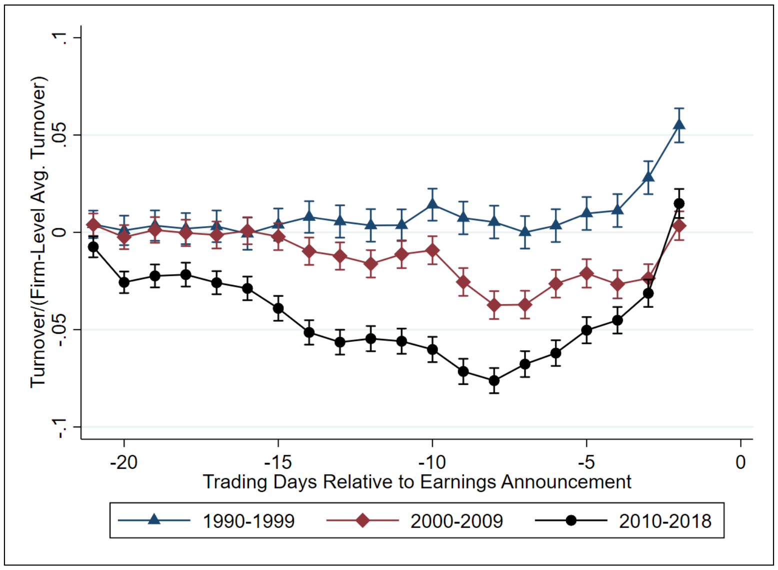

Figure E.3. Decline of pre-earnings turnover. Plot of β_τ estimated from the regression:

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

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

E.6 Quasi-experimental results

In this subsection, I replicate the results from the mechanisms section of the paper, but restrict to increases in passive ownership that arise from S&P 500 index addition and Russell 1000/2000 index rebalancing.

E.6.1 Pre-earnings turnover

Table E.6 shows that pre-earnings trading volume declines both after S&P 500 index addition and after switching from the Russell 1000 to the Russell 2000 index.

	Panel A: Binary Instrument			
	Russell		S&P	
	IV	RF	IV	RF
Post x Treated		-0.518		-0.607***
		(0.454)		(0.196)
Passive Ownership	-20.12***		-14.72*	
	(6.652)		(7.440)	
Observations	9,348	9,348	280,253	280,253
F-statistic	142.5		305.3	
	Panel B: Continious Instrument			
	Russell		S&P	
	IV	RF	IV	RF
Post x Treated		-8.227		-23.32***
x Passive Gap		(18.220)		(7.689)
Passive Ownership	-19.78***		-14.75**	
	(6.462)		(7.404)	
Observations	9,348	9,348	280,253	280,253
F-statistic	153.3		410	
Cross-sectional regression estimate	-11.49***	-11.49***	-11.49***	-11.49***

Table E.6 Effects of Russell rebalancing and S&P 500 index addition on pre-earnings volume. Estimates from:

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

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

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

Columns 1 and 3 are instrumental variables regressions. Columns 2 and 4 are reduced-form regressions. Panel A contains regressions from the binary treatment specification, while Panel B contains regressions from the continuous treatment specification. Columns 1 and 2 are from the Russell experiment, while columns 3 and 4 are from the S&P experiment. FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

E.6.2 Earnings responses

As a check on my cross-sectional regression results, I want to know whether firms which are added to the S&P 500 or switch from the Russell 1000 to Russell 2000 have larger earnings responses than the control firms. While the IV is a natural design for measuring the effect of passive ownership on price informativeness (see e.g., Appel et al. (2020)), it is not straightforward to map this to the earnings response regressions in the main body of the paper, as the coefficient of interest is on the interaction term between passive ownership and SUE. A more straightforward approach is a difference-in-differences, which interacts the treated and post × treated dummy variables with SUE.

Specifically, I run the following regression:

$$\begin{aligned} r_{i,t} = a + bSUE_{i,t} + cPost_{i,t} + dPost_{i,t} \times Treated_{i,t} + \\ ePost_{i,t} \times Treated_{i,t} \times SUE_{i,t} + FE + \epsilon_{i,t} \end{aligned} \tag{12}$$

where $r_{i,t}$ is the market-adjusted earnings-day return, in percentage points, Winsorized at the 1% and 99% level by year. FE are a set of cohort-by-firm fixed effects – which is why the standalone dummy variable for $Treated_{i,t}$ is not included. If being added to the S&P 500, or switching from the Russell 1000 to the Russell 2000 increases a firm's responsiveness to earnings news, e should be positive. I also construct a version of equation 12 which accounts for time-series changes in the treatment effect, replacing all instances of $Post_{i,t} \times Treated_{i,t}$ with $Post_{i,t} \times Treated_{i,t} \times PassiveGap_{i,t}$.

Table E.7 shows that earnings responses increase after S&P 500 index addition, and this effect is stronger for negative news. While the estimated coefficients are the same sign for firms which switch from the Russell 1000 to the Russell 2000, the results are statistically insignificant. This, however, might be due to the fact that the sample is much smaller. I omit the results decomposing earnings news into systematic/idiosyncratic and positive/negative components, as within any given firm-by-cohort set, there is no guarantee there will be at least one observation in each of these buckets.

E.6.3 Analyst coverage and accuracy

Table E.8 shows that analyst coverage increases for firms that get added to the S&P 500 index, but decreases for firms which switch from the Russell 1000 to the Russell 2000. For

	S&P Experiment				Russell Experiment			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
SUE	0.148***	0.148***			0.587***	0.593***		
	(0.021)	(0.021)			(0.083)	(0.077)		
$SUE \times 1_{Post=1} \times 1_{Treated=1}$	0.118***				0.177			
	(0.037)				(0.214)			
$SUE \times 1_{Post=1} \times PassiveGap$		6.797***				7.87		
		(1.829)				(8.094)		
$SUE \times 1_{SUE>0}$			0.231***	0.230***			0.796***	0.746***
			(0.038)	(0.038)			(0.151)	(0.158)
$\|SUE\| \times 1_{SUE\leq 0}$			-0.0851**	-0.0858**			-0.441***	-0.486***
			(0.034)	(0.034)			(0.095)	(0.085)
$SUE \times 1_{SUE>0} \times 1_{Post=1} \times 1_{Treated=1}$			0.0285				-0.397	
			(0.079)				(0.276)	
$\|SUE\| \times 1_{SUE\leq 0} \times 1_{Post=1} \times PassiveGap$			-0.189**				-0.642	
			(0.078)				(0.399)	
$SUE \times 1_{SUE>0} \times 1_{Post=1} \times 1_{Treated=1}$				3.904				-8.539
				(3.053)				(8.419)
$\|SUE\| \times 1_{SUE\leq 0} \times 1_{Post=1} \times PassiveGap$				-10.20**				-23.57
				(3.984)				(15.500)
Observations	256,946	256,946	256,946	256,946	8,658	8,658	8,658	8,658
R-squared	0.061	0.061	0.062	0.062	0.05	0.05	0.051	0.051

Table E.7 Effects of Russell rebalancing and S&P 500 index addition on earnings responses. Estimates from:

$$r_{i,t} = a + bSUE_{i,t} + cPost_{i,t} + dPost_{i,t} \times Treated_{i,t} + ePost_{i,t} \times Treated_{i,t} \times SUE_{i,t} + FE + \epsilon_{i,t}$$

where $r_{i,t}$ is the market-adjusted return on the effective earnings announcement date. Columns 1-4 are from the S&P experiment, while columns 5-8 are from the Russell experiment. The odd-numbered columns are from the binary treatment specification, while the even-numbered columns are from the continuous treatment specification FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

S&P 500 index additions, there is an increase in the dispersion of analyst estimates while there is no effect for the Russell experiment. Finally, there is a decrease in analyst accuracy both for the S&P 500 adds and the Russell 2000 switchers.

E.7 Ex-ante earnings uncertainty (IVD)

Table E.9 shows that the price of options which span earnings announcements increases both after S&P 500 index addition and after switching from the Russell 1000 to the Russell 2000 index.

	Panel A: Binary Instrument					
	Num. Est		SD(Est.)		Dist./SD(Est.)	
	S&P	Russell	S&P	Russell	S&P	Russell
Passive Ownership	112.9***	-12.69***	0.513***	-0.138	7.503***	3.814**
	(9.728)	(4.301)	(0.087)	(0.099)	(1.491)	(1.536)
Observations	277,009	9,233	264,762	8,738	212,775	8,423
F-statistic	309.9	145.5	312.1	153.6	273.8	148.6
	Panel B: Continious Instrument					
	Num. Est		SD(Est.)		Dist./SD(Est.)	
	S&P	Russell	S&P	Russell	S&P	Russell
Passive Ownership	111.4***	-13.11***	0.506***	-0.137	7.602***	3.891**
	(9.734)	(4.202)	(0.091)	(0.096)	(1.482)	(1.516)
Observations	277,009	9,233	264,762	8,738	212,775	8,423
F-statistic	423	156.4	432.9	164.3	375.5	158.5
Cross-sectional regression estimate	-11.64***	-11.64***	0.719***	0.719***	1.968***	1.968***

Table E.8 Effects of Russell rebalancing and S&P 500 index addition on analyst coverage and accuracy. Estimates from:

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

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

The odd-numbered columns in each panel are from the S&P experiment, while the even-numbered columns are from the Russell experiment. Panel A contains regressions from the binary treatment specification, while Panel B contains regressions from the continuous treatment specification. The outcomes are: (1) Num. Est. i.e., the number of analyst estimates, (2) SD(Est.) i.e., the standard deviation of analyst estimates and (3) Dist. i.e., the absolute distance between realized earnings per share and the mean estimate of earnings per share, FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are in parenthesis. All estimates are from instrumental variables specifications.

	Panel A: Binary Instrument			
	Russell		S&P	
	IV	RF	IV	RF
Post x Treated		0.0172**		0.00379
		(0.008)		(0.003)
Passive Ownership	0.319***		0.295***	
	(0.102)		(0.109)	
Observations	3,473	3,473	70,344	70,344
F-statistic	107.4		263.7	
	Panel B: Continious Instrument			
	Russell		S&P	
	IV	RF	IV	RF
Post x Treated		0.867**		0.021
x Passive Gap		(0.330)		(0.068)
Passive Ownership	0.326***		0.290***	
	(0.100)		(0.108)	
Observations	3,473	3,473	70,344	70,344
F-statistic	110.4		288.9	
Cross-sectional regression estimate	0.0958***	0.0958***	0.0958***	0.0958***

Table E.9 Effects of Russell rebalancing and S&P 500 index addition on IVD**.** Estimates from:

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

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

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

Columns 1 and 3 are instrumental variables regressions. Columns 2 and 4 are reduced-form regressions. Panel A contains regressions from the binary treatment specification, while Panel B contains regressions from the continuous treatment specification. Columns 1 and 2 are from the Russell experiment, while columns 3 and 4 are from the S&P experiment. FE are fixed effects for each firm-cohort group, as well as each month of index addition. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

References

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).

Azar, J., Schmalz, M. C., and Tecu, I. (2018). Anticompetitive Effects of Common Ownership. *Journal of Finance*, 73(4).

Bebchuk, L., Cohen, A., and Ferrell, A. (2009). What matters in corporate governance? *The Review of financial studies*, 22(2):783–827.

Beber, A. and Brandt, M. W. (2006). The effect of macroeconomic news on beliefs and preferences: Evidence from the options market. *Journal of Monetary Economics*, 53(8):1997–2039.

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

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

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).

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. (2008). The reduced form: A simple approach to inference with weak instruments. *Economics Letters*, 100(1):68–71.

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

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

Daniel, K., Klos, A., and Rottke, S. (2017). Overpriced winners. *Unpublished Working Paper, Columbia University.*

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.*

Hanson, S. G. and Sunderam, A. (2014). The growth and limits of arbitrage: Evidence from short interest. *The Review of Financial Studies*, 27(4):1238–1286.

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

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

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.

Manela, A. (2014). The value of diffusing information. *Journal of Financial Economics*, 111(1):181–199.

Massa, M., Schumacher, D., and Wang, Y. (2021). Who Is Afraid of BlackRock? *The Review of Financial Studies*, 34(4).

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

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

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