Online Appendix for
Passive Ownership and Price Informativeness

MARCO SAMMON*

August, 2022

A Stylized Facts

A.1 Visualizing the pre-earnings announcement drift

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

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

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

Figure A.1 plots these average cumulative market-adjusted returns by SUE decile for two different time periods: 2001-2007 and 2010-2018. The brown dashed line represents the average for firms with the most positive earnings surprises, while the blue dashed line represents the average for firms with the most negative earnings surprises. Consistent with

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

Ball and Brown (1968), in the left panel, the best news is preceded by positive market-adjusted returns while the worst news is preceded by negative market-adjusted returns.

Consistent with the trend in Figure 1 in the main body of the paper, firms in each decile move less before earnings days between 2010 and 2018 than between 2001 and 2007. The decline in pre-earnings drift is even stronger when comparing to the pre-2001 period, but that may be due to Regulation Fair Disclosure (Reg FD), implemented in August 2000, which limited firms' ability to selectively disclose earnings information before it was publicly announced.

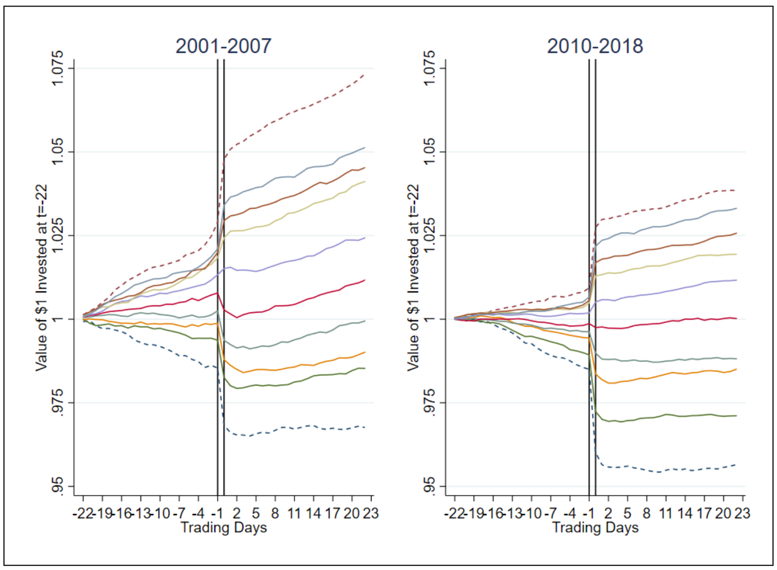

Figure A.1. Decline of pre-earnings drift by SUE decile. Each quarter, firms are sorted into deciles based on standardized unexpected earnings (SUE). Each line represents the cross-sectional average market-adjusted return of \$1 invested at $t = -22$. The brown dashed line represents the average for firms with the most positive earnings surprises, while the blue dashed line represents the average for firms with the most negative earnings surprises. The solid lines represent the averages for deciles 2 to 9.

A.2 Decomposition of earnings days' share of volatility

Figure A.2 decomposes the decline of QVS into the rise in volatility on earnings days and the decline in volatility on all other days (which determines the sum of squared returns from $t - 22$ to t i.e., the denominator of QVS). The trend in QVS was driven by a simultaneous increase in earnings day volatility (left panel) and a decrease in volatility on all other days (right panel).

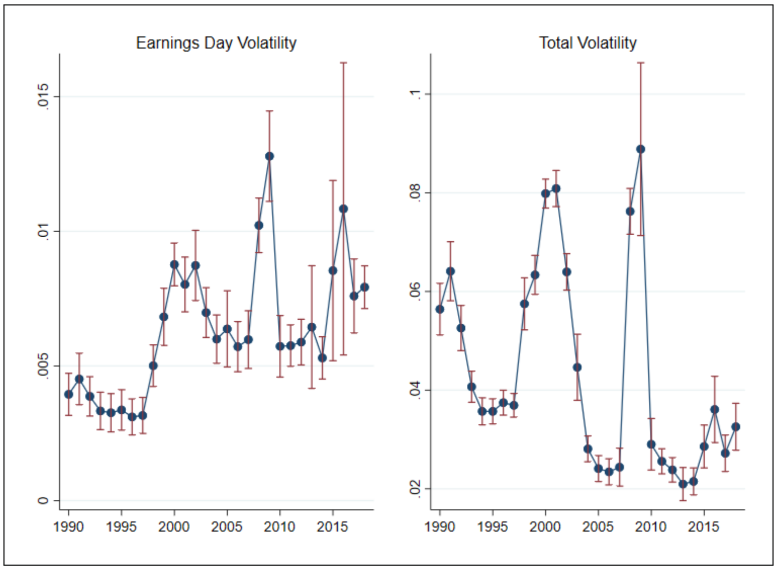

Figure A.2. Decomposition of QVS. This figure plots coefficients from a regression of the pieces of QVS on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient. The left panel has the squared earnings-day return on the left-hand side while the right panel has the sum of squared returns from $t - 22$ to t on the left-hand side. Standard errors represent 95% confidence intervals around the point-estimates. Standard errors are clustered at the firm level.

A.3 Relationship between DM and QVS

Given the similar time-series trends in QVS and DM, a natural question is whether they capture different information. By construction, they will both tend to be lower if the earnings-day return is large in absolute value. But, as discussed in the main body, they may sometimes yield different conclusions about pre-earnings announcement price informativeness because e.g., DM is sensitive to the level of volatility, while QVS is not. In terms of their statistical relationship, a univariate regression of QVS on DM has an R-squared of just under 50%. To visualize this relationship, Figure A.3 presents a scatter plot with QVS on the y-axis and DM on the x-axis.

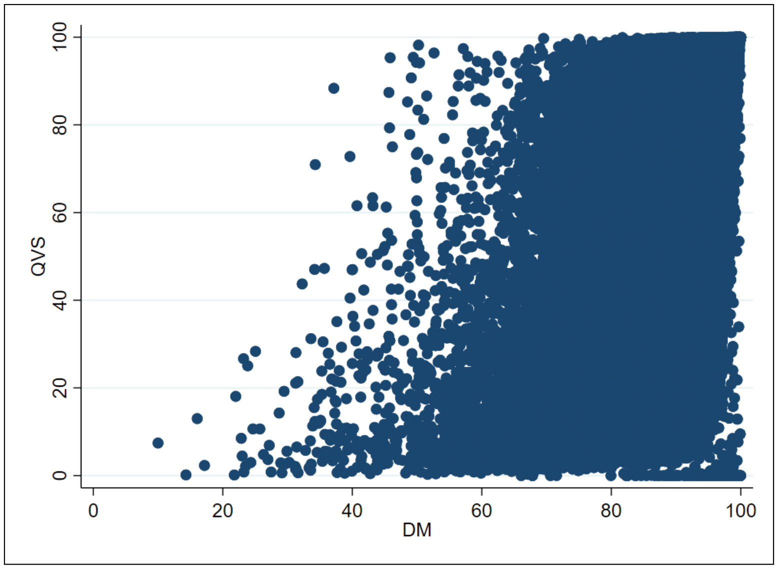

Figure A.3. QVS vs. DM. This figure is a scatter plot of QVS on DM. Each blue dot represents a single earnings announcement.

B Data details

B.1 Details on construction of control variables

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

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

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

Lagged book-to-market ratio: Book to market ratio of the stock at the end of the calendar month before the month of the earnings announcement from the WRDS financial ratios suite.

Total institutional ownership: The fraction of a stock's shares outstanding held by all 13-F filing institutions. Computed using the code here.

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

B.2 IBES

I merge CRSP to I/B/E/S (IBES) using the WRDS linking suite. Before 1998, nearly 90% of observations in IBES have an announcement time of "00:00:00", which implies the release time is missing. In 1998 this share drops to 23%, further drops to 2% in 1999, and continues to trend down to nearly 0% by 2015. This implies that before 1998, if the earnings release date was a trading day, I will always classify that day as the effective earnings date, even if earnings were released after markets closed, and it was not possible to trade on that information until the next trading day. This time-series variation in missing IBES release times is likely not driving my OLS estimates because in the main body of the paper, when ruling out the influence of Regulation Fair Disclosure, I show my results are quantitatively unchanged using only post-2000 data (i.e., the subsample where there are few missing earnings release times in IBES).

B.3 Computing passive and institutional ownership

To calculate passive ownership, I need to identify the holdings of passive funds, which I obtain from the Thompson S12 data.[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 unless the firm is also considered to have missing institutional ownership by this code (IO_MISSING = 1), in which case I also set passive ownership to missing. S12 data is only reported at the end of each calendar quarter, so to get a monthly estimate of passive ownership, I linearly interpolate passive ownership between quarter-ends. All results are quantitatively unchanged if I instead fix passive ownership at its last reported level between the ends of calendar quarters.

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

C Cross-Sectional Regressions

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

Although all the baseline regressions explicitly control for market capitalization, one might be worried that firm size has a non-linear effect on pre-earnings announcement price informativeness. To ameliorate this concern, Table C.1 replicates the baseline results, removing the control for market capitalization and instead including dummy variables for deciles of market capitalization, formed at the end of the previous calendar quarter. The results are quantitatively unchanged by including these fixed effects, suggesting that a non-linear effect of size on pre-earnings announcement price informativeness is not driving my OLS results.

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

Table C.1 Cross-sectional regression of price informativeness on passive ownership (fixed effects for deciles of market capitalization). Table with estimates of β from:
Price informativeness$_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where Price informativeness$_{i,t}$ is either $QVS_{i,t}$ or $DM_{i,t}$. Controls in $X_{i,t}$ include age, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All specification also include dummy variables for deciles of market capitalization, formed each quarter. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

C.2 Stocks with zero passive ownership

Previous studies on passive investing have excluded securities with zero passive ownership (Dannhauser, 2017). The logic is that there may be something special about the subset of securities passive funds avoid holding. Table C.2 contains the summary statistics and Table C.3 contains the baseline OLS regression results for the subsample with strictly positive passive ownership. Neither table is quantitatively changed from the corresponding one in the main body of the text by excluding these observations.

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

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

Table C.2 Summary Statistics (dropping observations with zero passive ownership). Cross-sectional equal-weighted means, standard deviations and distributions of price informativeness and passive ownership. Excludes all observations with zero passive ownership.

C.3 Comparison to previous work

While not identical, DM is similar to the price-jump measure of Weller (2018), which is also designed to capture the fraction of earnings information incorporated into prices before it was formally released. The difference is that DM uses gross returns, while price-jump uses

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

Table C.3 Cross-sectional regression of price informativeness on passive ownership (Dropping observations with zero passive ownership). Table with estimates of β from:
Price informativeness$_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where Price informativeness$_{i,t}$ is either $QVS_{i,t}$ or $DM_{i,t}$. Controls in $X_{i,t}$ include age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis. Excludes all observations with zero passive ownership.

net returns. To fix ideas, consider a net return version of DM: $\widehat{DM_{i,t}} = 100 \times r_{t-22,t-1}/r_{t-22,t}$. $\widehat{DM_{i,t}}$ solves one issue with $DM_{i,t}$ in that it is symmetric with respect to positive and negative returns.

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

DM and QVS are also related to absolute CARs around earnings announcements (Ball and Brown, 1968) and pre- and post- drug approval CARs (Manela, 2014). I believe that in my setting, as discussed in Weller (2018), DM and QVS have the advantage that they captures the share of information incorporated into prices before it is formally announced.

In the next two sections, I show that using Weller's price-jump measure or CARs does not change any of my empirical conclusions.

C.3.1 Relation to Weller (2018)

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

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

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

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

C.3.2 Relation to Manela (2014)

Manela (2014) examines the relationship between the value of information and how fast that information diffuses through financial markets around a different set of news events: FDA drug approvals. One of the quantities he uses to study this relationship is pre vs.

day-of vs. post announcement cumulative abnormal returns (CARs). The logic is that investors should trade more aggressively on faster-diffusing news. In equilibrium, this leads fast-diffusing news to have higher pre-announcement returns because of aggressive trading by informed insiders. On the other hand, slower diffusing news is mostly traded into prices after the announcement itself.

Although my setting is different, I can use the same logic to test whether passive ownership leads less information to be incorporated into prices before earnings announcements. By looking at drug approvals, Manela (2014) is focused on good news, so to create an analogue of these results for earnings announcements, I need to condition on the news itself. To this end, I split firms into 10 deciles based on their standardized unexpected earnings, and focus on firms in the top decile.[2] I then calculate the pre-announcement ($t = -5$ to $t = -1$), announcement-day ($t = 0$ to $t = 1$) and post-announcement ($t = 2$ to $t = 6$) cumulative market-adjusted returns (R_1, R_2 and R_3). If passive ownership decreases the amount of good news incorporated into prices ahead of time, more will be incorporated into prices on the day of, increasing R_2. An advantage of my measures relative to these CARs is that they ease comparison across stocks and time because DM and QVS capture the *share* of information incorporated into prices before it is formally announced.

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

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

$$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t} \tag{2}$$

where $Outcome_{i,t}$ is either $jump_{i,t}^{(a,b)}$ (with $b = 2$ and $a = 22$), R_1, R_2 or R_3. Controls in $X_{i,t}$ include time since listing (age), one-month lagged market capitalization, returns from month $t-12$ to $t-2$, one-month lagged book-to-market ratio and the institutional ownership ratio.

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

$X_{i,t}$ also includes CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility, all computed over the previous 252 trading days. Finally, the regression includes firm and year-quarter fixed effects. Standard errors are double-clustered at the firm and year-quarter level.

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

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

Figure A.1 suggests that the time-trend toward decreased DM was not equal for firms which ended up releasing good news and firms which ended up releasing bad news. To clarify this asymmetry, Figure C.1 presents a version of Figure A.1 which splits stocks into quintiles of SUE and quartiles of passive ownership using data between 2010 and 2018. Figure C.1 highlights two types of asymmetry. The first is that firms with low SUE have smaller pre-announcement drift than firms with high SUE. The second is that the effect of passive ownership on the pre-earnings drift is stronger for firms that end up releasing bad news than good news.

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

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^{5} b_j 1_{SUE_{i,t} \in Q_j} + \phi_t + \psi_i + e_{i,t} \tag{3}$$

where $1_{SUE_{i,t} \in Q_j}$ is an indicator variable equal to 1 if $SUE_{i,t}$ is in the j^{th} quintile of SUE in a given quarter (all results are similar running a version of Equation 3 with deciles of SUE instead of quintiles). Table C.5 contains the results. Column 1 shows that, consistent with Figure C.1, there is an unconditional asymmetry in DM between stocks which release

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

Table C.4 Passive ownership and alternative measures of pre-earnings announcement price informativeness. Estimates of β from:
$Outcome_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
$Outcome_{i,t}$ is either $jump^{(22,2)}_{i,t}$, the price jump measure from Weller (2018) with $a = 22$ and $b = 2$, $R_{1,i,t}$, the cumulative market-adjusted return from $t = -5$ to $t = -1$, $R_{2,i,t}$, the cumulative market-adjusted return from $t = 0$ to $t = 1$ or $R_{3,i,t}$, the cumulative market-adjusted return from $t = 2$ to $t = 6$ from Manela (2014). Controls in $X_{i,t}$ include age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so $0.01 = 1\%$ of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis. In Columns 2-4, the sample is restricted to firms in the top decile of SUE.

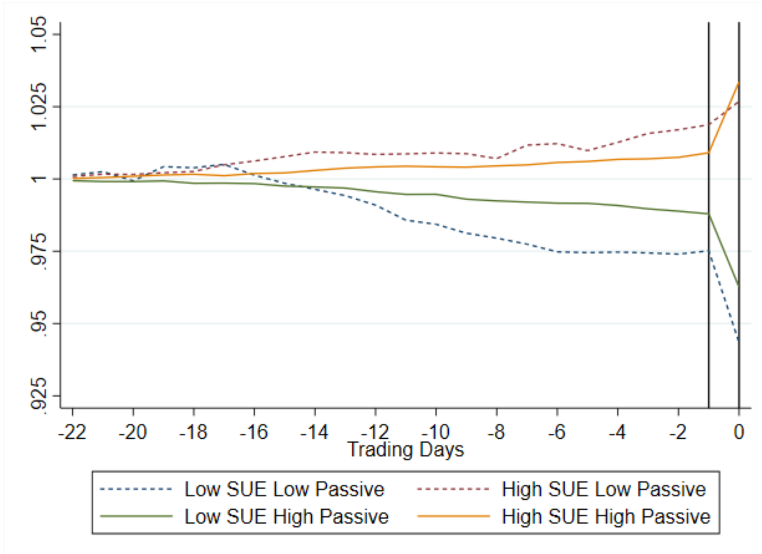

Figure C.1. Pre-earnings drift by SUE quintile and quartile of passive ownership. Each quarter, firms are sorted into quintiles based on standardized unexpected earnings (SUE) and quartiles based on their passive ownership share. Each line represents the cross-sectional average market-adjusted return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

good news and bad news. The last row of this column is the p-value from a test of $b_1 = b_5$, which suggests this difference is statistically significant. Column 3 shows the asymmetry is qualitatively unchanged by using value weights, instead of equal weights.

To quantify the role of passive ownership in this asymmetry (i.e., test for asymmetry within a given level of passive ownership), I run a second regression which includes interaction terms between the quintiles of SUE and passive ownership:

$$DM_{i,t} = \alpha + \beta Passive_{i,t} + \sum_{j=1}^{5} d_j 1_{SUE_{i,t} \in Q_j} + \sum_{j=1}^{5} c_j 1_{SUE_{i,t} \in Q_j} \times Passive_{i,t} + \phi_t + \psi_i + e_{i,t} \quad (4)$$

Column 2 of Table C.5 shows that there is an asymmetry between firms that release good and bad news for a given level of passive ownership, as c_1 is less than c_5. The last row of this Column is the p-value from a test of $c_1 = c_5$, which again suggests the difference is statistically significant. Column 4 shows this is also robust to using value weights instead of equal weights.

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

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

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

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

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

For every regression, the middle quintile is the omitted group. In columns 1 and 3, the last row of the table contains the p-value from a test of whether $b_1 = b_5$. In columns 2 and 4, the last row of the table contains the p-value from a test of whether $c_1 = c_5$. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

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

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

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

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

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

$$DM_{it}^n = 100 \times \begin{cases} \frac{1+r_{(t-22,t-1)}}{1+r_{(t-22,t+n)}} & \text{if } r_{(t,t+n)} > 0 \\ \frac{1+r_{(t-22,t+n)}}{1+r_{(t-22,t-1)}} & \text{if } r_{(t,t+n)} < 0 \end{cases} \tag{5}$$

Figure C.2 shows that the time-series trends in DM are similar for choices of n up to 5.

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

$$100 \times \sum_{\tau=-22}^{-1} r^2_{i,t+\tau} / \sum_{\tau=-22}^{n} r^2_{i,t+\tau} \tag{6}$$

Figure C.3 shows that the time-series trends in QVS are similar for choices of n up to 5.

These same concerns could also apply to the baseline OLS estimates. Table C.6 shows that including up to 5 days after the earnings announcement does not qualitatively change my baseline results for QVS or DM.

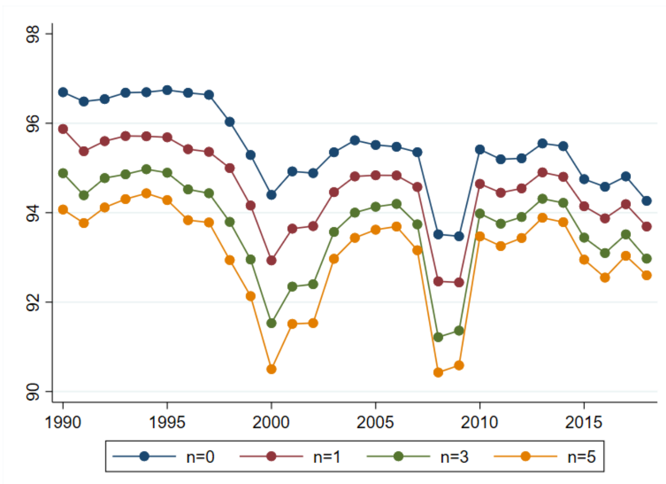

Figure C.2. Time series trends in DM^n. This figure plots coefficients from a regression of DM^n on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient.

C.6 Implied volatility difference

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

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

Table C.6 Sensitivity of QVS and DM results to including a n-day post-earnings-announcement window. Estimates of β from:

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

Where $Outcome^n$ is either $QVS^n_{i,t}$, a version of QVS that includes n days after the earnings announcement in the volatility attributed to the earnings announcement itself or $DM^n_{i,t}$, a version of DM that includes n days after the earnings announcement in the return attributed to the earnings day itself. Controls in $X_{i,t}$ include age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

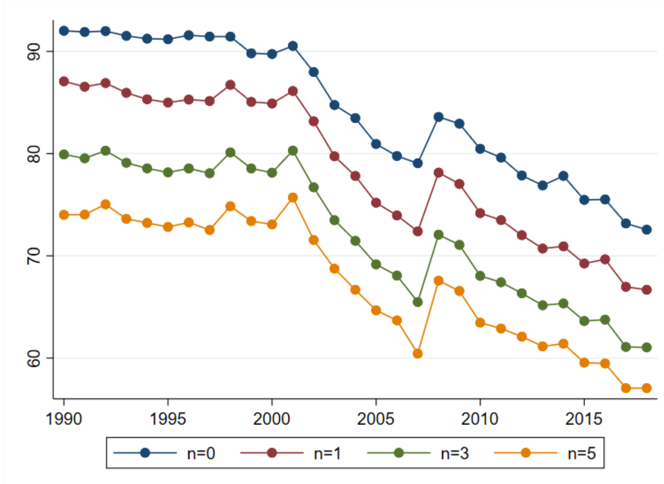

Figure C.3. Time series trends in QVS^n. This figure plots coefficients from a regression of QVS^n on a set of year dummy variables. The constant term for the omitted year (1989) is added to each coefficient.

volatility for short-maturity options (Beber and Brandt, 2006), b is selected so that it is at least 5 days after τ.[3]

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

$$\overline{IV}_{i,b} = Mean\left[IV_{i,(b-s,b),b} : b - s \in [\tau - 20, \tau - 1]\right] \tag{7}$$

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

$$IVD_{i,\tau} = \overline{IV}_{i,b} - \frac{1}{2}\left(\overline{IV}_{i,a} + \overline{IV}_{i,c}\right) \tag{8}$$

[3]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.

higher values of $IVD_{i,\tau}$ imply that options which span earnings announcements are relatively more expensive.[4] The units of IVD are percentage points of implied volatility.

Implied volatility is computed by OptionMetrics and runs from 1996 until the end of my sample. I use the WRDS linking suite to match the OptionMetrics data with CRSP. Following Kelly et al. (2016), I keep all options with positive open interest, and define at-the-money options as those with absolute values of delta between 0.4 to 0.5. For a firm/earnings-announcement pair to be included, it must be that a and b are no more than two months apart, and c is no more than one month after b.[5]

Figure C.4 plots the cross-sectional average of IVD by quarter. Numbers greater than zero are evidence that options which span earnings announcements are more expensive than those with surrounding maturities. Consistent with the increase in earnings-day volatility (i.e., the decline in QVS), on both an equal-weighted and value-weighted basis, IVD has increased by about 5 over the past 25 years.

C.7 Alternative explanations for the decline of price informativeness

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

C.7.1 Regulation Fair Disclosure (Reg FD)

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

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

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

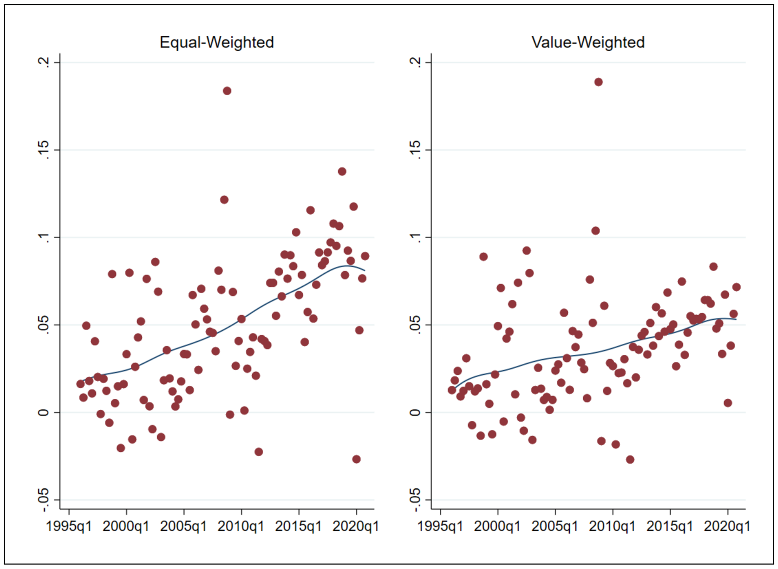

Figure C.4. Time-series trends in IVD**.** Equal-weighted and value-weighted averages of IVD by quarter. Red dots represent cross-sectional averages and blue lines represent LOWESS filters with bandwidths equal to 20% of quarters in the dataset.

information, and instead must release it to all investors at the same time.

Reg FD could be driving the trends in QVS and DM, as there was a large negative shock to the amount of information firms released before earnings announcements after it was passed. QVS and DM, however, continue to trend in the same direction after Reg FD was implemented. Reg FD could still explain these results if the value of information received by analysts before Reg FD decayed slowly. While this is possible, my prior is that information obtained in 2000 would not be relevant for more than a few years.

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

Another possibility is that Reg FD changed the way insiders (e.g., directors or senior officers) behaved, or led to changes in the enforcement of insider trading laws (Coffee, 2007).

If this were purely a time-series effect, however, it cannot be driving the OLS regressions which have time fixed-effects. To further rule out the insider behavior channel, I used the Thompson Insiders data to compute insider buys/sells as a percent of total shares outstanding for each firm in my dataset.

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

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

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

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

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

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

C.7.3 Effect of passive ownership on corporate governance

Given the literature on the effects of passive owners on corporate governance (Appel et al., 2016), one could worry that passive ownership's primary effect is to change governance, and then governance changes price informativeness. One mechanism would be that better governance leads to fewer information leaks, which in turn makes prices less informative before earnings announcements.[6]

To test this, I quantify corporate governance using the entrenchment index (E index) of Bebchuk et al. (2009). Using data from ISS between 1990 and 2018, I calculate this as the sum of indicator variables for the presence of: (1) a staggered (classified) board (2) a limitation on amending bylaws (3) a limitation on amending the corporate charter (4) a requirement of a supermajority to approve a merger (5) golden parachutes for management/board members and (6) a poison pill.[7] I then run a regression of the E index on passive ownership. Given that the E index is only defined annually, I use end of year data for passive ownership as well as all the control variables. Also, given that ISS coverage is not equally spread across the firm size distribution, I do not report the value-weighted regression results.

[6]There is, however, mixed evidence on the relationship between passive ownership and corporate givernance. 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.

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

	(1)	(2)	(3)
Passive Ownership	0.402	0.581	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.7 Passive ownership and entrenchment. Table with estimates of β from:
$EIndex_{i,t} = \alpha + \beta Passive_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + e_{i,t}$
where $EIndex$ is the entrenchment index of Bebchuk et al. (2009). Controls in $X_{i,t}$ include age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. Passive ownership is expressed as a decimal, so 0.01 = 1% of firm i's shares are owned by passive funds. Standard errors double clustered at the firm and year-quarter level in parenthesis.

Consistent with Gloßner (2018), Table C.7 shows there isn't a statistically significant relationship between governance and passive ownership. The effect of a 15% increase in passive ownership on the E index is less than 0.05, so the effect of passive ownership on governance is also economically small relative both to the mean (≈3) and the standard deviation (≈1.5). I also find that my baseline regressions are unchanged by explicitly controlling for the E index. Jointly, this evidence suggests that the relationship between passive ownership and corporate governance is not driving my results.

D Causal analysis

D.1 Firm size and passive ownership

As discussed in the introduction, a possible threat to identification is the relationship between passive ownership and firm size. Figure D.5 plots the relationship between the passive ownership share and the percentile of market capitalization for observations in December

2018. The relationship is positive with a univariate R-squared of 25%. For very large stocks (those in the top 20% of market capitalization) the relationship starts to break down and invert. One explanation for this is that mid-cap indices (e.g., the Russell 2000) have a relatively larger passive ownership share than large-cap indices (e.g. the Russell 1000) (Pavlova and Sikorskaya, 2022).

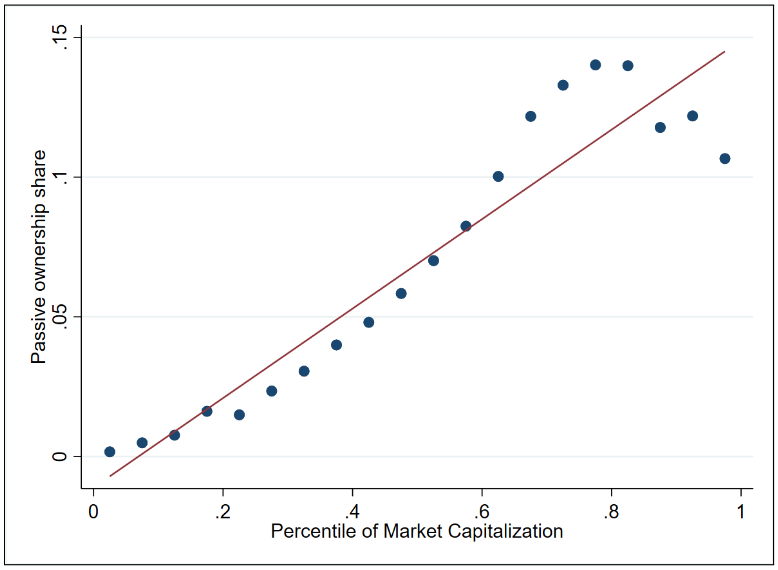

Figure D.1. Passive ownership and percentile of market capitalization. Data from 12/2018. Includes all firms with both non-missing passive ownership and non-missing market capitalization.

D.2 Russell Details

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

- Compute the number of shares outstanding/market capitalization on the index rebalancing date according to CRSP. To do this, start with the CRSP daily security file. Merge this with the list of dates from Ben-David et al. (2019) to identify the trading date closest to the Russell index rebalancing date.

- An adjustment has to be made if a PERMCO (permanent company identifier in CRSP) has multiple associated PERMNOs (permanent security identifier in CRSP). There are two broad cases to consider: (1) If only one of the PERMNOs is in the Russell 3000 universe, for each PERMNO, compute total market capitalization at the PERMCO level (2) If more than one of the PERMNOs is in the Russell 3000 universe, compute the market capitalization for each PERMNO individually.[8]

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

  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

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

release date of earnings information (RDQ), and the Russell index rebalancing date. To do this, start at RDQ, and apply all of the CRSP factor to adjust shares between RDQ and the rebalancing date.

- Map the Russell index member data to CRSP using the following procedure:
  - First, create a new CUSIP variable that is equal to historical CUSIP if that is not missing, and is equal to current CUSIP otherwise. Merge on this new CUSIP variable and date.
  - For the remaining unmatched firms, merge on ticker, exchange and date.
  - For the remaining unmatched firms that had non-missing historical CUSIP, but weren't matched on historical CUSIP to the Russell data, merge on current CUSIP and date.
  - For the remaining unmatched firms, merge on ticker and date. Note that in some of these observations, the wrong field is populated (e.g., the actual ticker was put into the CUSIP field in the Russell data), so that needs to be fixed before doing this last merge.
- Merge CRSP and Compustat using the CRSP/Compustat merged data.
- Use the following procedure to compute May market capitalization: If the shares outstanding from the Compustat data is larger than the shares outstanding from CRSP, use that number of shares outstanding to compute market capitalization. Otherwise, use the shares outstanding in the CRSP data to compute market capitalization. In either case, compute market capitalization using the closing price on the day closest to the index rebalancing date.

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

- Each May, rank stocks by market capitalization.
- Identify the 1000th ranked stock, and compute the bands as $\pm$ 2.5% of the total market capitalization of the Russell 3000.[9]

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

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

D.3 Alternative instruments

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

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

Again, following Coles et al. (2022), I choose the control firms to be those within ± 100 ranks of the upper band that were in the Russell 2000 the previous year. Figure D.2 shows the problem with this IV: the change in passive ownership associated with switching from the 2000 to the 1000 is small and temporary. Within 12 months of switching, passive ownership is almost back at the pre index-rebalancing level.

D.3.2 Blackrock's acquisition of Barclays Global Investors

Another possible instrument for passive ownership can be constructed around Blackrock's acquisition of Barclays' iShares ETF business in December 2009. This is not an ideal setting

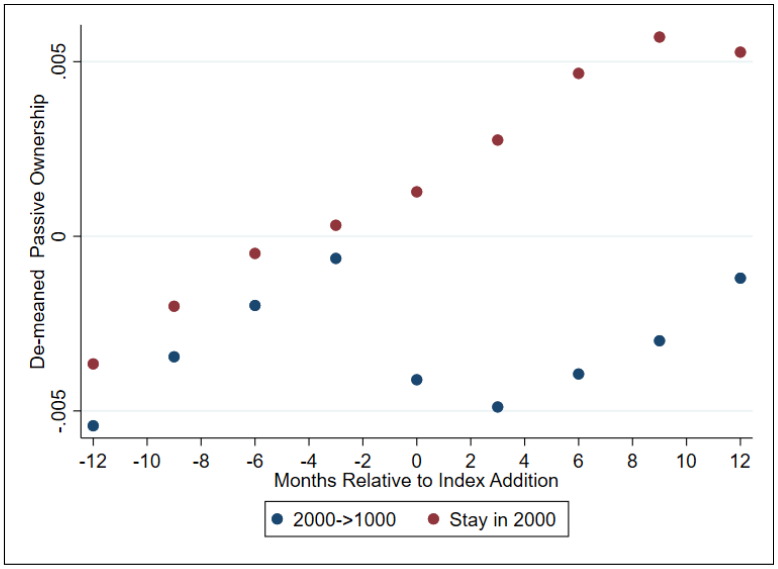

Figure D.2. Russell 1000/2000 Reconstitution and Changes in Passive Ownership. Average level of passive ownership for firms that stay in the Russell 2000 (control firms) and firms that moved from the Russell 2000 to the Russell 1000 (treated firms). Passive ownership is demeaned within each cohort.

for testing my hypothesis because: (1) My proposed mechanism has no predictions for the effects of increased concentration of ownership among passive investors (Azar et al. (2018), Massa et al. (2021)) and (2) While there may have been a *relative* increases in flows to iShares ETFs, compared to all other ETFs (Zou, 2018), I do not find a significant increase in overall ETF ownership for the stocks owned by iShares funds. Given that my right-hand side variable of interest is the percent of shares owned by passive funds, my proposed mechanism has no predictions for the effect of moving dollars from iShares ETFs to non-iShares ETFs.

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

In Table 4, the IV regressions are highly significant, while the reduced form regressions are insignificant. The concern is that, as discussed in Chernozhukov and Hansen (2008), a significant IV with an insignificant reduced form potentially indicates a weak instruments problem. 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., test whether the reduced form is insignificant.[10] At a high level, this is likely not a problem in my setting, as the first stage is very strong ($F > 300$ for the S&P experiment and $F > 200$ for the Russell experiment).

Further, as pointed out by Lochner and Moretti (2004), for a given IV standard error, the reduced form standard errors can be arbitrarily large or small. To formalize the argument, consider the case of a univariate structural regression and a single instrument. The model is

$$\text{Structural} : \quad y_i = \beta x_i + \varepsilon_i \tag{9}$$

$$\text{First stage} : \quad x_i = \gamma z_i + u_i \tag{10}$$

$$\text{Reduced Form} : \quad y_i = \overbrace{\beta\gamma}^{\alpha} z_i + \overbrace{\beta u_i + \varepsilon_i}^{v_i} \tag{11}$$

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

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

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

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

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

between x_i and ε_i:

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

Writing out the definition of the IV estimator:

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

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

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

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

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

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

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

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

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

And

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

Thus, their joint distribution is:

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

where

$$\rho_{\varepsilon,v} = \frac{E\left[v_i \varepsilon_i\right]}{\sqrt{\left(E\left[v_i^2\right]\right)\left(E\left[\varepsilon_i^2\right]\right)}} = \frac{\beta \frac{E[u_i \varepsilon_i]}{E\left[\varepsilon_i^2\right]} + 1}{\sqrt{\beta^2 \frac{E\left[u_i^2\right]}{E\left[\varepsilon_i^2\right]} + 2\beta \frac{E[u_i \varepsilon_i]}{E\left[\varepsilon_i^2\right]} + 1}} \tag{12}$$

Equation 12 implies that if the true $\beta = 0$, $\rho_{\varepsilon,v}$ will be equal to 1 and the t-statistics will be perfectly correlated asymptotically. Alternatively, if $\beta \neq 0$, then $\rho_{\varepsilon,v}$ will be less than 1 and these two t-statistics will not be perfectly correlated, even asymptotically. Thus, it is possible to have a significant IV estimate and insignificant reduced form estimate and this becomes more likely as $\rho_{\varepsilon,v}$ decreases.

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

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

and taking their ratio yields:

$$\frac{t_{\widehat{\beta}_{IV}}}{t_{\widehat{\alpha}_{RF}}} = \frac{\sqrt{\left(N^{-1} \sum_i \widehat{v}_i^2\right)}}{\sqrt{\left(N^{-1} \sum_i \widehat{\varepsilon}_i^2\right)}} \xrightarrow{p} \sqrt{\frac{E\left[v_i^2\right]}{E\left[\varepsilon_i^2\right]}} = \sqrt{\frac{\beta^2 E\left[u_i^2\right] + 2\beta E\left[u_i \varepsilon_i\right] + E\left[\varepsilon_i^2\right]}{E\left[\varepsilon_i^2\right]}} \tag{13}$$

Thus, under the $\beta = \alpha = 0$ null, these t-statistics will be perfectly correlated asymptotically.[11]

Empirically, in Table 4, I am testing whether $\widehat{\beta}_{IV} = 0$ and $\widehat{\alpha}_{RF} = 0$. With this in mind,

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

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

1. Equation 13 shows that even fixing the IV t-statistic, the standard errors in the reduced form can be arbitrarily large or small depending on the correlation between the residuals in the structural and first stage regressions. Suppose, for example, $t_{\widehat{\beta}_{IV}} = -2.5$ so the IV is statistically significant, the true β is -0.5, and $E\left[u_i^2\right] = E\left[\varepsilon_i^2\right] = 1$. Then, if the covariance between u_i and ε_i is lower than -0.4, the reduced form coefficient will not be both negative and significant at the 5% level. This is because with these parameters $\sqrt{\frac{\beta^2 E\left[u_i^2\right] + 2\beta E[u_i\varepsilon_i] + E\left[\varepsilon_i^2\right]}{E\left[\varepsilon_i^2\right]}} = 1.28$ and $-2.5/1.28 > -1.96$.
2. More generally, in my setting I expect $\beta < 0$ i.e., passive ownership decreases 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 13, increasing the ratio of the IV t-statistic to the reduced form t-statistic.
3. As the number of observation in my sample grows, we expect both the IV and reduced form t-statistics to increase because this will decrease:
   $\hat{\sigma}_\varepsilon^2 = (y_i - \hat{\beta}_{IV}x_i)'(y_i - \hat{\beta}_{IV}x_i)/N$
   and
   $\hat{\sigma}_v^2 = (y_i - \hat{\alpha}_{RF}z_i)'(y_i - \hat{\alpha}_{RF}z_i)/N$

Economic mechanism for correlation in error terms: Shadow indexing

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

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

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

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

Now, in my IV, I measure the average difference in passive ownership for firms around the cutoff before index rebalancing to estimate the change in passive ownership a firm will receive from being added to the index, which I call $PassiveGap_{i,t}$. But suppose that firm i also gets added to several sub-indices, so the true increase in passive ownership is larger than $PassiveGap_{i,t}$. Recalling the first stage regression:

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

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

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

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

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

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

where $\tilde{\varepsilon}_{i,t} = \varepsilon_{i,t} + \beta \cdot shadow_{i,t}$. In this setting $u_{i,t}$ and $\tilde{\varepsilon}_{i,t}$ are going to have negative covariance, because $shadow_{i,t}$ is positively related to $passive_{i,t}$. And if $\beta < 0$, then $\beta E\left[u_i \varepsilon_i\right] > 0$, which according to Equation 13 would tend to make the reduced form have a smaller t-statistic than the IV.

Simulation evidence

I use simulations to understand just how large the correlation between u_i and ε_i would need to be to generate a scenario where I fail to reject the null via the reduced form but reject the null via the IV. Specifically, I simulate the setup in Equation 9 (the model with a univariate structural regression and a single instrument), varying the sign and the strength of the correlation between u_i and ε_i. Given that the sample size matters (both $\hat{\sigma}_\varepsilon^2$ and $\hat{\sigma}_v^2$ depend on N), I choose $N = 30,000$ to match the number of observations in Panel A of Table 4. I set $\beta = -0.25$, $\gamma = 0.5$, although all results are similar using any $\beta < 0$ and $\gamma \neq 0$. Finally, to ensure that the IV and reduced form estimates are not statistically significant in every simulation, I add additional noise to the system, scaling all ε by 5 and all u by 10.

Figure D.3 plots the fraction of simulations where the t-statistic from the IV is less than -1.96, but the t-statistic from the RF is greater than -1.96. The first dot on the far left of the plot shows that even if u_i and ε_i are uncorrelated, the RF is less likely to be statistically significant than the IV. This is not surprising, as even if $E[u_i \varepsilon_i] = 0$, the ratio in Equation 13 will be bigger than one.

The blue dots show that as the correlation between u_i and ε increases, the RF becomes more statistically significant on average. This is because the numerator in Equation 13 shrinks, as this positive covariance between u_i and ε is being multiplied by β, which is less than 0. Finally, the red dots show that as the correlation between u_i and ε becomes more negative, the RF becomes even less significant on average than the IV. In this case, the negative covariance between u_i and ε is being multiplied by the negative beta, which increases the numerator of Equation 13.

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.

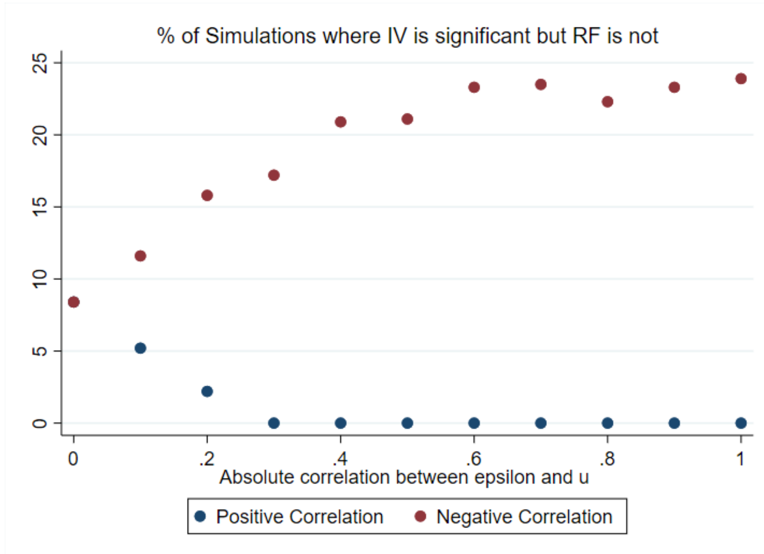

Figure D.3. Comparison of statistical significance. Each dot represents the percentage of simulations where the instrumental variables specification is statistically significant, but the reduced form is not. The blue dots are from simulations where ε positively correlated with u, while the red dots are from simulations where ε is negatively correlated with u. Moving from left to right increases the (absolute) correlation between ε and u.

D.5 Effect of treatment on total institutional ownership

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

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

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

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

I have two additional pieces of evidence to address the concern that total institutional ownership, rather than passive ownership, is driving my results: In the cross-sectional OLS regressions, I can and do explicitly control for total institutional ownership. In fact, I find there is significant cross-sectional variation in passive ownership within various levels of institutional ownership. For example, Figure D.4 plots passive ownership against institutional ownership in 12/2018. These two quantities are positively correlated, with a univariate R-squared of about 50%. This high correlation, however, is to be expected because passive ownership is included in total institutional ownership.

The second piece of evidence is in Table D.1, where I replicate all the instrumental variables and reduced form regressions, including total institutional ownership on the right hand side. All the results are quantitatively unchanged from Table 4 in the main body of the paper.

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

Table D.1 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness (conditioning on total institutional ownership). Estimates from:

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

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

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

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

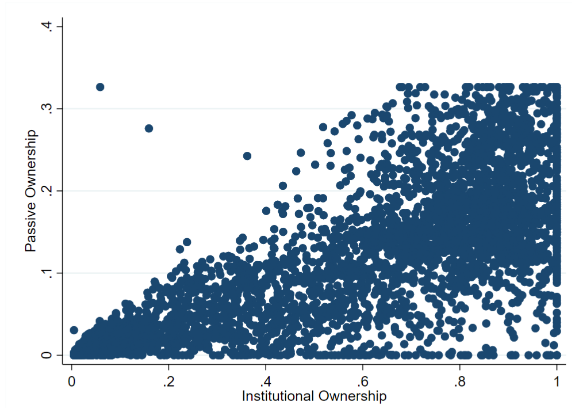

Figure D.4. Passive Ownership vs. Institutional Ownership. Plot of passive ownership against total institutional ownership in 12/2018. Both quantities are Winsorized at the 1% and 99% level.

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

To further allay concerns about my original IV, I have implemented a separate design that uses a different form of identifying variation. My original IV uses time-series variation because it compares differences in informativeness before and after index switching. Alternatively, I leverage cross-sectional variation in index changes that are generated by broad market movements in a small window just before index membership is decided. Specifically, I build on the logic in Bernstein (2015) and use the interaction between a firm's CAPM beta at the end of March and the cumulative market return from the start of April to the Russell ranking date in May to instrument for passive ownership from July to the following March. For example, in 2010, I use the interaction of CAPM beta and the cumulative market return from April 1st to May 28th to instrument for passive ownership between 7/2010 and 3/2011.

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

IVs in the main body of the paper (cross-sectional vs. time series).

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

Column 1 of Table D.2 shows the first stage. The relationship between the instrument and passive ownership is positive, consistent with the positive relationship between passive ownership and firm size documented in Figure D.5.[12] Further, the F statistic of 26 suggests the instrument is not weak (Stock and Yogo, 2002). Columns 2 and 4 are the IV regressions, which show point estimates similar in magnitude to the IVs in the main body of the paper. Columns 3 and 5 present the reduced form regressions, which are also negative and statistically significant.

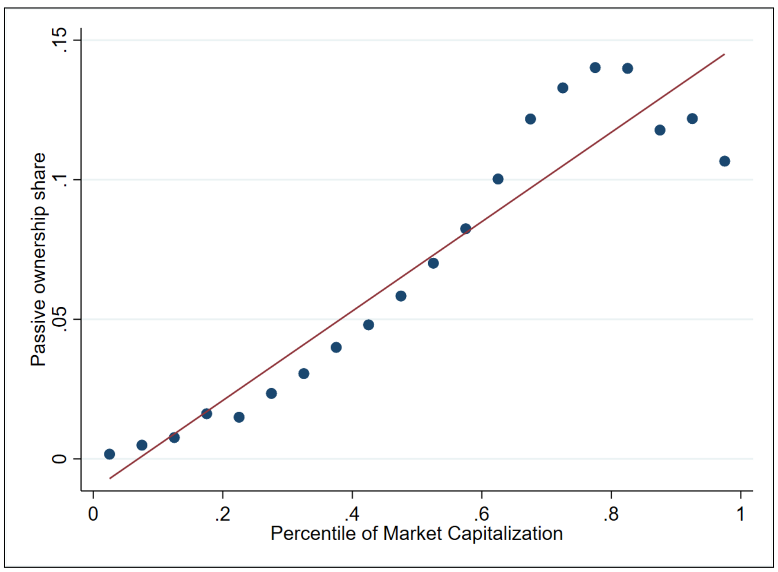

Figure D.5. Passive ownership and percentile of market capitalization. Data from 12/2018. Includes all firms with both non-missing passive ownership and non-missing market capitalization.

[12]Figure D.5 also shows that this relationship between passive ownership and market capitalization breaks down for very large firms. To this end, in this IV design I exclude firms that were in the S&P 500 at the end of March, but results are quantitatively unchanged by including these observations.

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

Table D.2 IV estimates for effect of passive ownership on pre-earnings announcement price informativeness (alternative instrument). Estimates from:

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

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

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

where $Outcome_{i,t}$ is QVS or DM, $\beta_{i,March(t)}$ is firm is CAPM beta at the end of March and $r_{m,April(t)\to May(t)}$ is the market return from the start of April to the Russell ranking date in May. Column 1 is a first-stage regression. Columns 2 and 4 are instrumental variables regressions. Columns 3 and 5 are reduced-form regressions. FE are fixed effects for the interaction between dummy variables for deciles of market capitalization, formed at the end of March, and dummy variables for each year. Standard errors, double clustered at the firm and quarter level, are in parenthesis.

E Mechanisms Details

E.1 Trends in Earnings Responses

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

$$100 \times r_{i,t} = \alpha + \beta SUE_{i,t} + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \tag{14}$$

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

Motivated by the asymmetries documented in Figure C.1 and Table C.5, I also design an earnings-response regression which allows for different reactions to positive and negative surprises:

$$100 \times r_{i,t} = \alpha + \beta_p 1_{SUE_{i,t} \geq 0} \times SUE_{i,t} + \beta_n 1_{SUE_{i,t} < 0} \times |SUE_{i,t}| + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \tag{15}$$

I run regressions 14 and 15 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.

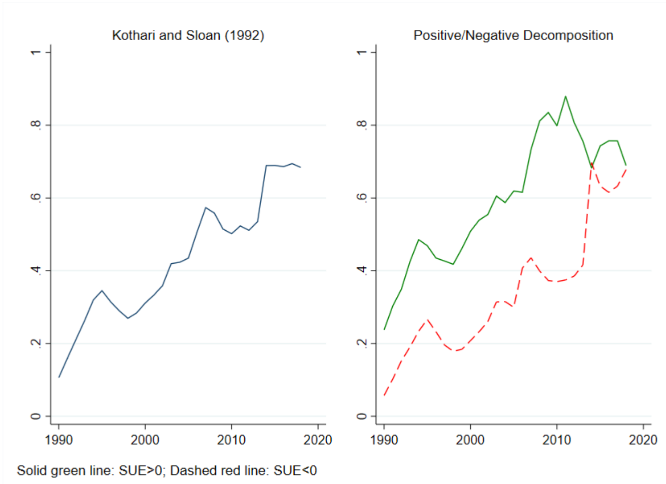

Figure E.1. Trends in Earnings Response. Left panel has estimates of β from:

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

run in 5-year rolling windows. Right panel has estimates of β_1 and β_2 from Equation 15 i.e., breaking SUE into positive and negative components. Controls in $X_{i,t}$ include age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All regressions contain year-quarter fixed effects, ϕ_t and firm fixed effects ψ_i.

E.2 Robustness of earnings response results

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

$$\begin{aligned} 100 \times r_{i,(t,t+n)} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} + \\ \gamma_1 \left(Sys.SUE_{i,t} \times Passive_{i,t}\right) \gamma_2 \left(Idio.SUE_{i,t} \times Passive_{i,t}\right) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t} \end{aligned} \tag{16}$$

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

Table E.1 shows that even including up to 5 days in $r_{i,(t,t+n)}$ does not change that high passive stocks are especially responsive to idiosyncratic news and that this is robust to using value weights or equal weights.

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

Table E.1 Sensitivity of earnings-response regressions to including a n-day post-earnings-announcement window. Estimates from:
$100 \times r_{i,(t,t+n)} = \alpha + \beta_1 SUE_{i,t} + \phi_1 Passive_{i,t} +$
$\gamma_1 \left(Sys.SUE_{i,t} \times Passive_{i,t}\right) \gamma_2 \left(Idio.SUE_{i,t} \times Passive_{i,t}\right) + \gamma X_{i,t} + \phi_t + \psi_i + \epsilon_{i,t}$
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 age, one-month lagged market capitalization, returns from t-12 to t-2, one-month lagged book-to-market ratio, total institutional ownership, CAPM beta, CAPM R-squared, total volatility and idiosyncratic volatility. All Columns contain year-quarter fixed effects and firm fixed effect. Standard errors double clustered at the firm and year-quarter level in parenthesis.

E.3 Trends in pre-earnings turnover

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

$$AT_{i,t+\tau} = \alpha + \sum_{\tau=-21}^{22} \beta_\tau \mathbf{1}_{\{i,t+\tau\}} + e_{i,t+\tau} \tag{17}$$

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

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

E.4 CRSP volume vs. total volume

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

E.5 IV estimates for mechanisms regressions

In Table E.2, I reproduce all the regressions from the mechanisms section of the paper using the IVs built on Russell 1000 to 2000 switchers and S&P 500 index additions.

	Panel A: Russell								
	IVD	Bloomberg	Downloads	CAT	Num Est	SD Est	Dist/SD(Est)	Updates	Time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Passive Ownership	0.281***	-4.47***	13.02***	-16.09**	-5.235	0.708*	5.876***	0.253	-0.35
	(0.098)	(1.092)	(2.879)	(7.363)	(4.552)	(0.399)	(1.669)	(0.348)	(1.459)

	Panel B: S&P								
	IVD	Bloomberg	Downloads	CAT	Num Est	SD Est	Dist/SD(Est)	Updates	Time
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Passive Ownership	0.335**	-1.1	24.57***	-14.72	111.5***	0.592***	6.561***	1.577***	-13.02***
	(0.139)	(0.805)	(2.133)	(10.424)	(11.220)	(0.137)	(1.792)	(0.324)	(4.246)

	Panel C: Earnings Responses	
	Russell	S&P
	(1)	(2)
$Idio.SUE \times Post \times$ $Treated \times PassiveGap$	18.22**	6.093**
	(8.565)	(2.508)

Table E.2 IV Estimates for mechanism regressions

For Panels A and B, estimates are from:

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

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

For Panel C, estimates are from:

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

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

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

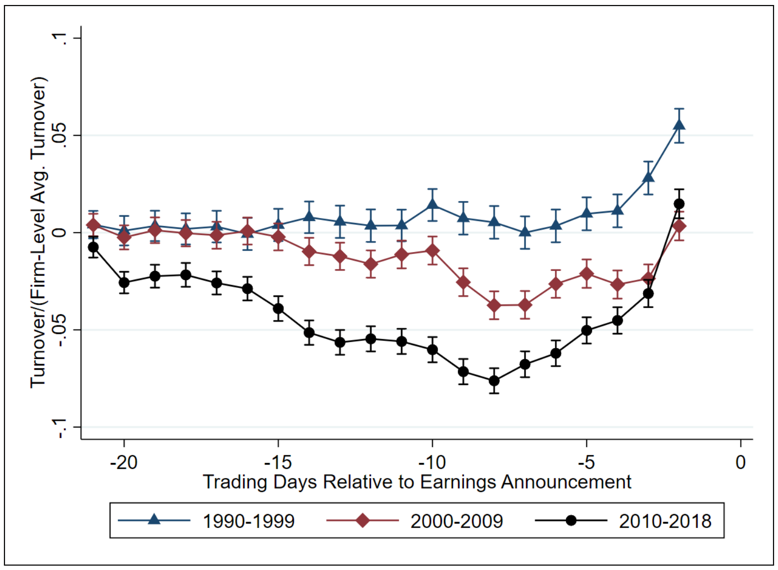

Figure E.2. Decline of pre-earnings turnover. Plot of β_τ estimated from the regression:

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

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.

References

Aghion, P., Van Reenen, J., and Zingales, L. (2013). Innovation and institutional ownership. *American economic review*, 103(1):277–304.

Appel, I., Gormley, T. A., and Keim, D. B. (2020). Identification using russell 1000/2000 index assignments: A discussion of methodologies. *Critical Finance Review, Forthcoming.*

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

Asquith, P., Pathak, P. A., and Ritter, J. R. (2005). Short interest, institutional ownership, and stock returns. *Journal of Financial Economics*, 78(2):243–276.

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

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

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

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

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

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

Bernstein, S. (2015). Does going public affect innovation? *The Journal of finance*, 70(4):1365–1403.

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

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.

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

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

Coles, J. L., Heath, D., and Ringgenberg, M. C. (2022). On index investing. *Journal of Financial Economics*, 145(3):665–683.

Dannhauser, C. D. (2017). The impact of innovation: Evidence from corporate bond exchange-traded funds (ETFs). *Journal of Financial Economics*, 125(3).

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

Gloßner, S. (2019). Investor horizons, long-term blockholders, and corporate social responsibility. *Journal of Banking and Finance*, 103.

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*.

O'Brien, P. C. and Bhushan, R. (1990). Analyst following and institutional ownership. *Journal of Accounting Research*, 28:55–76.

Pavlova, A. and Sikorskaya, T. (2022). Benchmarking intensity.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2012). The short of it: Investor sentiment and anomalies. *Journal of Financial Economics*, 104(2):288–302.

Stambaugh, R. F., Yu, J., and Yuan, Y. (2015). Arbitrage asymmetry and the idiosyncratic volatility puzzle. *The Journal of Finance*, 70(5):1903–1948.

Stock, J. H. and Yogo, M. (2002). Testing for weak instruments in linear iv regression.

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

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.