Passive Investing and Market Quality*

Philipp Höfler **Christian Schlag** **Maik Schmeling**

November 19, 2024

*First version: June 15, 2023. We thank Andrew Ellul, Monika Gehde-Trapp, Valentin Haddad, Holger Kraft, Mamdouh Medhat, Loriana Pelizzon, John Shim, Julian Thimme, Mihail Velikov, Christian Wagner, as well as conference and seminar participants at the SFS Cavalcade 2024, BI Oslo, and Goethe University for helpful comments and suggestions. We are grateful for research funding by the German Science Foundation (DFG) as part of the Research Unit on *Intermediary Asset Pricing*.

Höfler: Goethe University Frankfurt. Email: hoefler@finance.uni-frankfurt.de. Schlag: Goethe University Frankfurt and Leibniz Institute for Financial Research SAFE. Email: schlag@finance.uni-frankfurt.de. Webpage. Schmeling: Goethe University Frankfurt and Centre for Economic Policy Research (CEPR), London. Email: schmeling@finance.uni-frankfurt.de. Webpage.

Passive Investing and Market Quality

Abstract

We show that an increase in passive exchange-traded fund (ETF) ownership leads to stronger and more persistent return reversals. Exploiting exogenous changes due to index reconstitutions, we further show that more passive ownership causes higher bid-ask spreads, more exposure to aggregate liquidity shocks, more idiosyncratic volatility and higher tail risk. We examine potential drivers of these results and show that higher passive ETF ownership reduces the importance of firm-specific information for returns but increases the importance of transitory noise and a firm's exposure to market-wide sentiment shocks.

JEL Classification: G12, G14

Keywords: passive investing, ETFs, market efficiency, reversal

"ETPs may also affect the values of the underlying securities and the overall quality of financial markets – a concern that both industry and academic studies have recently expressed."

`SEC-NYU Dialogue on Exchange-Traded Products`[1]

1. Introduction

The importance of passive investing has risen substantially over the last decade, largely driven by the widespread adoption of passive exchange-traded funds (ETFs). This shift has raised concerns by policymakers and regulators that markets may have become more fragile and less efficient since ETFs may deteriorate liquidity and foster short-term speculation. A key concern is that passive investing crowds out investors that trade on fundamental information, which implies that markets may have become slower or less effective in processing and impounding such information into prices.

In a nutshell, the skeptical view of passive investing argues that more capital allocated to passive products, e.g. due to lower fees of passive vehicles, can make markets less efficient (e.g. Gârleanu and Pedersen, 2022), since it crowds out sophisticated, active investors that trade on news about fundamentals. For example, Jappelli (2023) shows that passive allocation mandates lead to persistent price pressure that grows with market prices and creates a structured form of noise. In the model of Bond and Garcia (2022), investors that switch to passive index products are relatively uninformed. As a consequence, more passive investing lowers price efficiency of the index but also decreases liquidity of individual stocks and introduces noise in prices of firms included in the index.

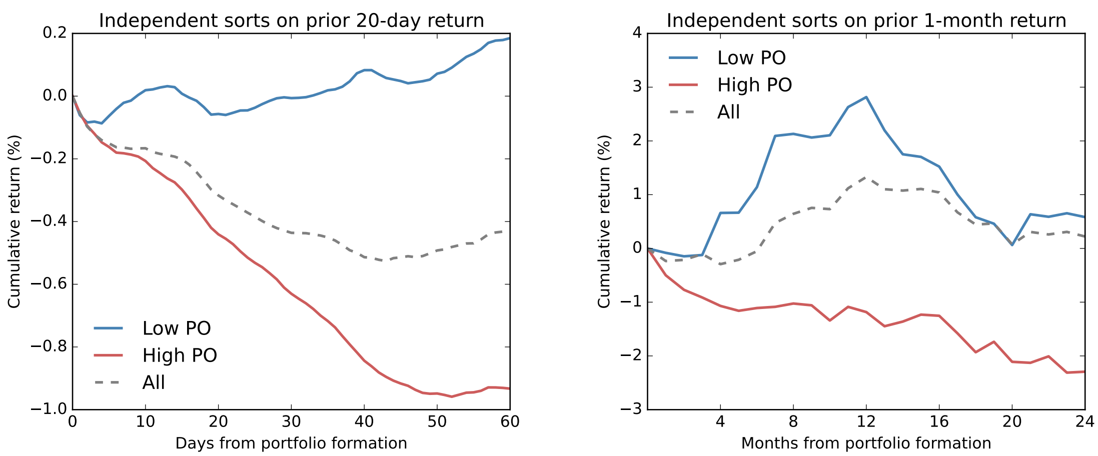
However, passive ETFs might also increase efficiency precisely because they lower transaction costs, facilitate short selling, and thus speed up the process of incorporating new information into prices. To date, the literature on this question is inconclusive. Against this backdrop, and guided by recent theoretical work, we are interested in how passive ETF ownership affects *market quality*, i.e. the liquidity and price efficiency of a market (e.g. O'Hara and Ye, 2011).

As a starting point to understand the impact of passive ETF ownership (PO) on liquidity, we test how PO affects short-term return reversals. Short-term reversal strategies go long (short) stocks with low (high) returns over a recent period and trade on the tendency of

[1] See `www.sec.gov/news/speech/speech-piwowar-2017-09-08`.

stock returns to revert after liquidity-induced price bounces (Jegadeesh, 1990). The size and speed of reversals are thus informative about the ability of market makers to satisfy liquidity needs. To examine the link between PO and return reversal, we run independent double sorts (value weighted) on past one month returns and PO using all U.S. stocks in the CRSP/Compustat sample. We find that stocks with high passive ETF ownership have much stronger reversal than those with low passive ETF ownership, which lends credence to the view that PO reduces stock liquidity and price informativeness.

Figure 1 summarizes the results for stocks in the high and low PO quintile. The economic size of the effect of PO on return reversals is large. For example, in double sorts based on the prior 20-day return with daily rebalancing (Panel (a)), stocks in the high PO quintile fall by about 1% over the subsequent 50 trading days before the effect levels out. In contrast, stocks with low PO initially fall by about ten basis points and quickly recover afterwards. In double sorts with monthly rebalancing, high PO stocks show a persistent reversal pattern of up to minus 2% over 24 months (Panel (b)).



(a) Daily returns

(b) Monthly returns

Figure 1: Winner-minus-loser performance by passive ETF ownership. This figure shows average cumulative post-formation excess returns of winner-minus-loser portfolios, which are independently sorted into quintiles based on prior month's return and passive ETF ownership. Portfolios are value weighted and prior return quintiles based on NYSE breakpoints. Panel a (b) shows cumulative daily (monthly) returns during 60 days (24 months) after portfolio formation. The sample spans all U.S. based common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

To alleviate concerns about endogeneity when testing for a link between passive ETF ownership and reversals, we make use of index reconstitutions in the Russell 1000/2000 indices to obtain exogeneous variation in passive ETF ownership following the setup in Chang, Hong, and Liskovich (2015), its application to ETFs in Ben-David, Franzoni, and Moussawi (2018), and the extension in Appel, Gormley, and Keim (2019), which allows us to cover a much longer sample period relative to the two aforementioned studies. Intuitively, index reconstitutions lead to shifts in passive ownership by ETFs that track the underlying index because they have to adjust their holdings if a stock switches index membership. We use this exogenous variation to study how changes in PO affect the exposure of a stock to the short-term reversal factor in the year after index reconstitution. Corroborating our results described above, we find a strong and significant effect of PO on stocks' exposures to short-term reversal factor returns.

The increase in return reversals and exposure to the reversal factor strongly suggest that changes in PO affect stock liquidity, as suggested by the theoretical work of Bond and Garcia (2022). Going beyond their model, we further test this conjecture by looking at other dimensions of liquidity and show that higher passive ETF ownership increases bid-ask spreads, liquidity risk (Pástor and Stambaugh, 2003), and idiosyncratic volatility. We again run these regressions using exogenous variation in PO due to index reconstitutions, such that they cannot easily be explained by reverse causality. Our findings are threefold. First, we look at the effect of PO on bid-ask spreads and contribute to the ongoing debate on how liquidity is shaped by noise trading (Peress and Schmidt, 2020). Specifically, through the lens of passive ETFs in general, and identification via index reconstitutions in particular, we find that an increase in noise trading lowers liquidity, supporting the predictions of inventory risk models (e.g. Ho and Stoll, 1981; Grossman and Miller, 1988). Second, we show that PO increases a stock's liquidity beta, consistent with the argument that ETFs create correlated demand for liquidity of their underlying stocks (Agarwal, Hanouna, Moussawi, and Stahel, 2018). Last, we show that PO increases idiosyncratic volatility, which has important implications for stock liquidity. More precisely, Jiang, Vayanos, and Zheng (2020) argue that higher idiosyncratic volatility discourages investors from absorbing demand shocks in stocks with high PO.

Having documented a significant effect of passive ETF ownership on stock liquidity, a natural question is whether passive investors affect asset prices directly. Recent theoretical work supports the notion that passive investors decrease the aggregate demand elasticity of an asset, leading to more price impact of flows (Haddad, Huebner, and Loualiche, 2021). Our results of lower liquidity and more exposure to aggregate liquidity shocks of stocks with more

PO are well in line with this finding, and suggest that stocks with more PO should be more prone to extreme price moves in response to flows (also see, e.g. Brunnermeier and Pedersen, 2009, on liquidity spirals). To isolate such tail risk, we make use of the option-implied tail risk measures proposed in Bollerslev and Todorov (2011). These measures capture jump risk in returns but are not affected by diffusive volatility and thus capture a different dimension of risk than pure return volatility as studied in Ben-David et al. (2018). Moreover, using option-based tail risk circumvents the problem that extreme events are rare and hard to measure in the data.

Using variation in passive ETF ownership due to index reconstitutions, we show that higher PO indeed causes a significant increase in both left and right tail jump risk. The economic size of this effect is such that a one standard deviation increase in PO raises left (right) tail risk by about 19.2 (19.4) percentage points. The magnitude of the effect is large relative to the average left (right) tail risk of 26% (23%). This finding suggests that changes in PO do indeed affect asset price dynamics beyond simple return volatility and that (option) markets price the effects of changes in passive ETF ownership.

Our findings so far are broadly in line with recent theoretical work on the effects of indexing as discussed above and, among other things, confirm earlier empirical work supporting the notion that passive ETFs decrease liquidity and increase non-fundamental noise (Ben-David et al., 2018; Brown, Davies, and Ringgenberg, 2021). We now turn to the second dimension of market quality and test whether PO affects price efficiency. Specificially, we do so by testing for the relative importance of firm fundamentals in driving stock prices (e.g. Bond and Garcia, 2022; Jappelli, 2023).

To do so, we start by decomposing stock return variance into market-wide news, firm-specific (public and private) news, and return noise following the methodology developed in Brogaard, Nguyen, Putnins, and Wu (2022).[2] We then test how PO affects the relative importance of different types of information using exogenous variation in PO due to index reconstitutions. Our results show that higher PO significantly increases the importance of noise but that it reduces the importance of firm-specific information.[3] More specifically, a one

[2]We explain this decomposition, which is based on structural vector autoregressions and a decomposition into transitory and permanent return responses to shocks, in more detail later in the paper.

[3]This finding echoes earlier results in the literature, e.g., by Brown et al. (2021) who use portfolio sorts to test this prediction of their model. We contribute to this literature by providing a decomposition of returns into different drivers and then testing the effect of PO on different types of news in a causal framework.

standard deviation increase in passive ETF ownership (≈ 7 percentage points relative to an average of 9%), raises the variance share of noise by about 6 percentage points but decreases the share of firm-specific information by 9 percentage points. Interestingly, the large drop in firm-specific information is fully accounted for by firm-specific *private* information. In other words, passive ownership does not seem to affect the importance of public news but that it decreases the importance of private news for stock prices.

Since firm-specific news are relatively less important for stock prices with high PO compared to noise, we test whether a higher PO increases the exposure of stocks to market-wide sentiment shocks and whether changes in PO are related to a stock's mispricing (Stambaugh, Yu, and Yuan, 2012). We find evidence for both effects, i.e. stocks with higher PO react more strongly to sentiment shocks and have, on average, higher mispricing. Additional tests demonstrate that this finding is unlikely to be explained by the phenomenon that new ETFs are often launched in popular stocks in which investor sentiment is currently high (Ben-David, Franzoni, Kim, and Moussawi, 2023).

Taken together, our results suggest that the decrease in liquidity that comes with more passive ETF ownership stems from an increase in short-term noise trading, which decreases the importance of firm-specific news for stock returns but amplifies exposure to transitory, market-wide shocks that are unrelated to fundamentals.

We provide a number of additional results and robustness checks and show that also including passive mutual fund ownership does not significantly change our main results, that results based on the index reconstitution experiment are robust to various parameter choices (e.g. using different window sizes around the index cutoff), and that our results for the effect of PO on tail risk are robust to using a simple, return-based measure that does not require option data.

Related literature. We build on a growing literature that studies the effects of passive ownership on asset prices and market outcomes. An important paper in the context of our study is Ben-David et al. (2018) which shows empirically that ETF ownership increases return volatility, which they attribute to an increase in return noise, and that ETF flow-induced returns tend to be reversed. Relative to their findings, we contribute to this literature by showing that *passive ETF ownership* significantly increases returns to standard short-term reversal strategies especially for the liquidity-related component of reversals. Notably, we show that reversals of stocks with high PO are large (up to 2%) and very persistent (up to two years). Going beyond Ben-David et al. (2018) we make use of Russell index reconstitutions to identify the causal effect of passive ETF ownership for a much longer sample period based on the approach in Appel et al. (2019) and document a significant effect of PO on stock liquidity and option-implied tail risk.[4] Finally, we provide causal evidence that ETF ownership reduces (increases) the importance of firm-specific information (transitory noise) in driving returns.

Haddad et al. (2021) show that the massive increase of passive investors led to significantly lower aggregate demand elasticity and Pavlova and Sikorskaya (2023) document that index membership attracts inelastic demand that lowers the risk premium over a prolonged period. We contribute to this literature on the effects of passive investing on market efficiency by providing causal evidence that higher PO decreases the importance of firm-specific information while simultaneously increasing the importance of return noise. Our results on the increased exposure to sentiment shocks, lower liquidity, and higher tail risk further enhance our understanding of the asset pricing implications on passive investing. Having said that, a growing literature also documents no detrimental effects (e.g. Coles, Heath, and Ringgenberg, 2022) or beneficial effects of (passive) ETF ownership on market quality, see, e.g. Glosten, Nallareddy, and Zou (2021); Ahn and Patatoukas (2022); Filippou, He, and Zhou (2022). Our findings on the effect of PO on market quality and the result that stock prices incorporate less firm-specific news but more transitory noise can speak to these results as well. Moreover, since our paper provides evidence on the *causal* effect of PO on stock return drivers in a unified setting, we complement findings in the earlier literature based on correlations or lead-lag relationships.

2. Data and descriptive statistics

Our analysis focuses on U.S. equity ETFs that aim to physically replicate an index. The construction of our ETF sample is similar to Ben-David et al. (2018); Ben-David, Franzoni, and Moussawi (2019). First, we identify all ETFs in the Thomson Reuters Global Ownership database using the SecClsCode="ETF" condition and link each ETF to its CUSIP using information from OWNSECMAP. Next, we restrict all ETFs to Lipper Objective Codes for

[4] Agarwal et al. (2018) document that ETF ownership increases commonality in liquidity. We contribute to this literature by using a much longer sample period.

U.S. equity and omit active or leveraged ETFs based on their fund names.[5] The final sample consists of 872 ETFs from June 1997 to December 2021.

We rely on information from Thomson Reuters Global Ownership to calculate passive ETF ownership in stock i as percent of its total market capitalization. Formally,

$$\mathrm{PO}_{i,t} = \sum_{j=1}^{J} \frac{\mathrm{ValueHeld}_{i,j,t}}{\mathrm{MktCap}_{i,t}} \times 100, \tag{1}$$

where $\mathrm{ValueHeld}_{i,j,t}$ is the dollar amount of ETF j invested in stock i, obtained from the most recent quarterly investment report filed to the SEC. In between quarter ends we hold ETF ownership constant. Appel, Gormley, and Keim (2016) point out that funds were only required to report their holdings to the SEC twice per year until May 2004. To account for missing or unreported holdings, we extrapolate ownership of ETF j in stock i for a maximum of three quarters. If a stock is not held by any passive ETF we set PO to zero.

Our stock universe covers the CRSP sample, i.e., all U.S. common stocks (share codes 10 or 11) listed on NYSE, AMEX and NASDAQ (exchange codes 1, 31, 2, 32, 3, 33). We do not exclude financial firms. Firm characteristics are obtained from Chen and Zimmermann (2022). Detailed descriptions of the construction of these variables are shown in Internet Appendix IA.1.

Figure 2 shows the evolution of average passive ETF stock ownership over time and labels key events in the rise of passive ETFs. Over the course of our sample period, average passive ETF ownership has seen a steep rise to more than 10%, i.e. on average about 10% of stock ownership is in the hands of passive ETF at the end of 2021.

Table I provides an overview of our stock-month sample. In each month, we sort all stocks into quintiles based on their passive ETF ownership and report average firm characteristics. By construction, average passive ETF ownership increases monotonically from the low PO quintile to the high PO quintile. Whereas firms with low market capitalization are located in the low PO quintile, this is not the case for large firms. The average size of the third PO

[5] Following Ben-David et al. (2018), we use the following Lipper Objective Codes obtained from the CRSP Mutual Fund Database to identify U.S. equity funds: CA, EI, G, GI, MC, MR, SG, SP, BM, CG, CS, FS, H, ID, NR, RE, TK, TL, S, and UT. To filter out active leveraged ETFs, the keywords are: active, momentum, smart, alpha, factor, long/short, long short, arbitrage, bull, bear, bullbear, ultra, short, inverse, and leveraged.

Table I: Average stock characteristics by PO quintiles. This table presents average stock characteristics across passive ETF ownership quintiles. Quintile portfolios are formed each month based on the share of a stock's market capitalization held by passive ETFs. In Panel (a), equal weighted averages are taken across time and across firms. Panel (b) shows time-series averages of periodic value-weighted cross-sectional averages. Size measures a stock's market capitalization in million USD, turnover the trading volume divided by the number of shares outstanding, and volatility the standard deviation of daily returns within a month. CBOP/AT is cash-based operating profits divided by total assets. Asset growth is the percentage change in total assets. Momentum is the cumulative stock returns over the past 12 months, leaving out the most recent month. Data are monthly and cover all U.S. common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

(a) Equal weighted summary statistics

	Low PO	2	3	4	High PO
PO (%)	0.29	2.40	4.35	6.44	9.14
Size (Million USD)	448.85	1941.52	13060.19	6324.59	3612.71
Turnover (%)	17.77	14.95	17.68	20.82	21.12
Volatility (%)	5.02	3.66	2.96	2.75	2.77
B/M	0.95	0.75	0.57	0.56	0.64
CBOP/AT	0.03	0.07	0.13	0.15	0.15
Asset growth	18.99	25.80	22.50	19.13	13.18
Momentum	13.51	19.31	17.83	14.27	11.46

(b) Average value weighted cross-sectional summary statistics

	Low PO	2	3	4	High PO
PO (%)	0.54	2.75	4.72	6.31	9.10
Size (Million USD)	20431.76	94770.11	157744.44	43427.00	27241.65
Turnover (%)	19.18	16.18	15.26	18.76	20.92
Volatility (%)	2.66	2.26	2.00	2.10	2.23
B/M	0.59	0.44	0.38	0.42	0.49
CBOP/AT	0.14	0.16	0.20	0.19	0.17
Asset growth	26.00	37.84	20.44	17.98	14.83
Momentum	31.70	30.50	26.54	22.74	21.96

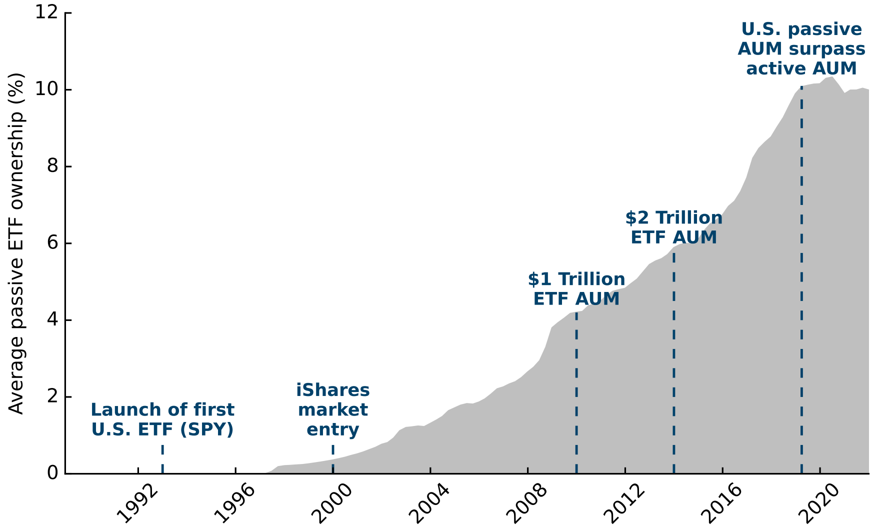


Figure 2: Growth of passive ETF ownership. This figure shows the (cross-sectional) average fraction of a stock's total market capitalization held by passive ETFs that focus on broad-based U.S. equity (shaded area). Construction details of this variable are described in Section 2. Data are at the quarterly frequency and cover all U.S.-based common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

quintile exceeds the average size of the high PO quintile by a factor of more than three. For example, Apple Inc. is in the third PO ownership quintile for most of our sample period. Furthermore, we observe that firms with high ETF ownership exhibit higher turnover and lower volatility. Note that these statistics cannot be interpreted causally as passive ETF ownership may be endogenous.[6]

3. Passive ownership and market quality

In this Section, we show that passive ETF ownership increases the duration and size of return reversals (Section 3.1), that it decreases stock liquidity (Section 3.2), and the consequences of these effects, i.e. that higher passive ETF ownership increases a stock's tail risk (Section 3.3).

[6]For example, Ben-David et al. (2018) show that ETFs cause an increase in stock return volatility.

3.1. *Return reversal*

We start with the impact of changes in passive ETF ownership on liquidity, which is one of the key concerns of policy makers and market regulators.[7] To do so, Table II expands on the evidence in Figure 1 and presents results for independent double sorts based on passive ETF ownership quintiles and different reversal signals. Passive ETF ownership and reversal (one month returns) are measured over the month prior to portfolio formation and we report value weighted returns. Panel (a) of this table shows that winner-minus-loser returns (row "High-Low") decrease monotonically when moving from the first quintile of stocks with low PO to the fifth quintile with high PO. This pattern is economically sizeable: Whereas low passive ETF stocks have average winner-minus-loser returns of 0.0% per month, stocks with high passive ETF ownership show winner-minus-loser returns of -0.5% per month.

Since we argue that our findings are driven by the impact of passive ETF ownership on market making capacity through higher noise in returns, we can further sharpen these tests. Specifically, we want to compute reversal returns after removing return components that are unrelated to liquidity. Prior research has shown that industry returns are associated with short-term momentum (Moskowitz and Grinblatt, 1999) and announcement returns with the post-earnings-announcement drift (Ball and Brown, 1968). We thus subtract the industry return and/or the cumulative return around a stock's most recent earnings announcement from prior stock returns as suggested by Dai, Medhat, Novy-Marx, and Rizova (2024) who study the returns to liquidity provision and market making capacity.

Table II, Panel (b), reports average excess returns for announcement-adjusted reversal (REVX), Panel (c) for industry-relative reversal (IRR), and Panel (d) for announcement-adjusted industry-relative reversal (IRRX). Across all panels, reversals are most pronounced among stocks in the high passive ETF quintile. However, short-term reversals strongly increase as we correct prior returns for industry momentum and post-earnings-announcement drift (PEAD). For example, buying recent winners and selling recent losers based on IRRX generates average excess returns of -0.77% per month ($t = -3.43$) for stocks with high PO

[7]For example, the IMF Global Financial Stability Report (10/2022) argues that: *"The growing role of passive investing that offers daily redemptions to retail investors, coupled with signs of increased herding and concentration, has made market liquidity more vulnerable to rapid changes in sentiment. Moreover, the ability of arbitrageurs such as hedge funds to take advantage of temporary price dislocations in asset markets, and therefore act as liquidity providers, may be limited."*

Table II: Double sorts on prior return and passive ETF ownership. This table shows the results of a bivariate independent-sort portfolio analysis of prior returns and passive ETF ownership. At the end of each month, all stocks are sorted into quintiles based on previous month's (adjusted) return and into quintiles based on the percentage of their total market capitalization held by passive ETFs. The table reports value-weighted average portfolio returns. t-statistics (in parentheses) are adjusted for heteroscedasticity and autocorrelation using 5 lags following Newey and West (1987). The sample covers all U.S. based common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

(a) Standard reversal (REV)

	Low PO	2	3	4	High PO
Low REV	0.45	0.97	0.79	0.73	0.95
2	0.97	0.90	0.84	0.87	0.91
3	0.54	1.18	0.70	0.86	0.79
4	0.98	0.76	0.76	0.84	0.72
High REV	0.45	0.90	0.65	0.56	0.45
High-Low	0.00	-0.07	-0.14	-0.17	-0.50
t-statistic	(0.0)	(-0.21)	(-0.36)	(-0.54)	(-2.09)

(b) Announcement-adjusted reversal (REVX)

	Low PO	2	3	4	High PO
Low REVX	0.52	0.96	0.95	0.87	1.09
2	0.92	1.02	0.85	0.73	0.83
3	0.70	0.94	0.83	0.96	0.76
4	0.98	0.83	0.65	0.90	0.75
High REVX	0.30	0.79	0.60	0.48	0.42
High-Low	-0.22	-0.17	-0.35	-0.39	-0.66
t-statistic	(-0.55)	(-0.5)	(-0.98)	(-1.23)	(-2.65)

(c) Industry-relative reversal (IRR)

	Low PO	2	3	4	High PO
Low IRR	0.50	0.98	1.09	0.72	0.98
2	0.54	1.05	0.96	0.94	0.82
3	0.73	1.08	0.63	0.75	0.84
4	0.74	0.77	0.67	0.63	0.61
High IRR	0.45	0.68	0.79	0.70	0.44
High-Low	-0.04	-0.31	-0.30	-0.02	-0.54
t-statistic	(-0.11)	(-1.1)	(-0.85)	(-0.08)	(-2.59)

(d) Ann.-adjusted industry-relative reversal (IRRX)

	Low PO	2	3	4	High PO
Low IRRX	0.63	1.02	1.09	0.91	1.07
2	0.67	0.94	1.19	0.82	0.85
3	0.76	0.95	0.49	0.80	0.75
4	0.74	0.80	0.69	0.58	0.70
High IRRX	0.44	0.64	0.63	0.65	0.31
High-Low	-0.20	-0.38	-0.46	-0.26	-0.77
t-statistic	(-0.56)	(-1.32)	(-1.43)	(-0.97)	(-3.43)

but only -0.20% per month ($t = -0.56$) among stocks with low PO.

To control for other characteristics that might drive our findings, we also run cross-sectional regressions with additional control variables. Specifically, we regress monthly stock excess return on previous month's PO, prior return (REV), and their interaction. We use weighted least squares regressions with market capitalization as weight to alleviate the impact of very small firms (Hou, Xue, and Zhang, 2020). Furthermore, we add further interaction terms to ensure our findings are not a result of any correlation between passive ETF ownership and turnover, size, or volatility.

Table IA.3 in the Internet Appendix presents average slopes and t-statistics (adjusted for heteroscedasticity and autocorrelation following Newey and West (1987)). While REV alone is not a significant predictor of future excess returns (model 1), we find a strong impact of passive ETF ownership on the size of this effect (model 2). Specifically, moving up one PO quintile decreases the profitability of a winner-minus-loser strategy by 17 bps ($t = -4$). In contrast, we also document that prior returns positively predict future returns among stocks with high turnover, consistent with the short-term momentum documented by Medhat and Schmeling (2022). However, controlling for turnover $\times$ PO does not does not drive out the effect of passive ETF ownership on reversal. Apart from this effect, all other interaction terms lose their significance when included alongside with $REV \times PO$ (model 10).[8]

3.2. *Identifying the effect of passive ownership on liquidity*

Our results based on double sorts and cross-sectional regressions are suggestive of a detrimental effect of PO on stock market liquidity, which is in line with Bond and Garcia (2022). To better understand this result, we further inspect the effect of PO on various other dimensions of liquidity at the stock-level. The first measure is the bid-ask spread, which could be affected by PO through an increase in noise trading. The relation between stock liquidity and noise trading is an important, yet open question in the literature (Peress and Schmidt, 2020).[9] Our paper contributes to this debate by looking at shifts in noise trading through

[8]We also run cross-sectional regressions with interaction terms to test for the importance of industry- and/or PEAD-adjusted returns and document these results in Table IA.4 in the Internet Appendix. We find that our main conclusions drawn from the double sorts in Table II are borne out in these cross-sectional regressions as well.

[9]In particular, adverse selection models support the view that noise trading improves liquidity (e.g. Glosten and Milgrom, 1985; Kyle, 1985), whereas inventory risk models predict a deterioration in liquidity (e.g. Ho

the lens of passive ETFs in general, and identification via index reconstitutions in particular. The second measure is a stock's sensitivity to market-wide liquidity shocks. Agarwal et al. (2018) argue that common ETF ownership creates correlated demand for the liquidity of constituent stocks. Since passive ETFs approximately hold the market portfolio, we expect an increase in individual stock betas to market-wide liquidity shocks. The third variable is a stock's short-term reversal beta, which we use to confirm our previous results at the stock-level and to test the predictions of Bond and Garcia (2022). Finally, we look at idiosyncratic volatility, which is also commonly used as a gauge for mispricing and noise trading. Based on the model model of Jiang et al. (2020), we expect passive funds to increase idiosyncratic volatility.

In order to alleviate concerns about endogeneity of passive ETF ownership and liquidity, we now provide complementary evidence based on index reconstitutions in the Russell 1000/2000 indices. Specifically, we follow the setup pioneered by Chang et al. (2015), Ben-David et al. (2018) and the extension in Appel et al. (2019). Intuitively, index reconstitutions lead to shifts in passive ownership by ETFs that track the underlying index and have to adjust their holdings if a stock switches index membership. Since this setup is well known in the literature, we relegate details about the mechanics of index reconstitutions, the construction of additional controls, as well as results for the first-stage regression to instrument passive ETF ownership to Appendix A.1. Based on instrumented PO from index reconstitutions, we now test for the causal effect of passive ETF ownership on different liquidity proxies and report results in Table III.

The first proxy is a stock's bid-ask spread, measured as the average percentage spread in the twelve months after index reconstitution. We report results for regressions of bid-ask spreads on instrumented PO in the first column of Table III. We find that a one-standard deviation increase in passive ETF ownership significantly increases average bid-ask spreads in the year after index reconstitution by 0.9 of one standard deviation.

To assess another dimension of liquidity, we ask whether higher PO also impacts on the exposure of stocks to *market-wide* liquidity shocks. To answer this, we measure a stock's sensitivity to aggregate market liquidity by a regression of monthly excess returns on the market factor and the Pástor and Stambaugh liquidity factor (separately for each stock and

and Stoll, 1981; Grossman and Miller, 1988).

each twelve month period after index reconstitution).[10] Using the estimated slope coefficient from this regression as dependent variable in the second stage of the index experiment, we find a positive and highly significant impact of passive ETF ownership on stocks' liquidity risk. Specifically, a one standard deviation increase in ETF ownership results in a 0.9 standard deviation increase in liquidity betas as reported in the second column of Table III.

We next turn to the effect of PO on return reversal. While we cannot directly test how PO affects short-term reversal returns at the firm-level, we compute firm-level stock return *exposures* to the short-term reversal factor. We do so by running regressions of stock returns on the short-term reversal factor using daily data in the twelve months *after* index reconstitution separately for each firm and twelve month period. The third column in Table III reports results for regressions of short-term reversal exposures on instrumented passive ETF ownership. We find that a one-standard deviation increase in passive ETF ownership leads to a statistically and economically significant increase in the exposure to the short-term reversal factor by 0.7 standard deviations in the cross-section of firms.

As a final exercise to gauge the effect of PO on liquidity, we test whether PO increases idiosyncratic volatility.[11] The final column of Table III reports regression results and shows that higher PO indeed significantly increases idiosyncratic volatility ($t = 2.79$). The economic size of this effect is such that a one standard deviation change in passive ETF ownership increases idiosyncratic volatility by about 0.6 of a standard deviation. This increase in idiosyncratic volatility suggests a reduced market-making capacity and complements findings in Ben-David et al. (2018) for total stock return volatility.

3.3. *Passive ownership and tail risk*

As a final exercise, we ask whether markets *price* the effects of passive ETF ownership on stock liquidity. Intuitively, lower liquidity coupled with a higher exposure to aggregate liquidity shocks suggests that stock returns should be more prone to extreme price moves (see, e.g. Brunnermeier and Pedersen, 2009, on liquidity spirals). Recent theoretical work also supports the idea that a higher share of passive investors leads to a greater price impact of flows.

[10] Since there are only 12 observations per firm and year, we obviously cannot use a sophisticated factor model.

[11] The computation of idiosyncratic vol follows Ang, Hodrick, Xing, and Zhang (2006) and is constructed from regressions of daily stock returns on the Fama-French three factor model in the twelve months *after* index reconstitutions, i.e. we have one estimate per firm and year.

Table III: Impact of passive ETF ownership on liquidity, reversal, and idiosyncratic volatility. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on liquidity, reversal, and idiosyncratic volatility. Specifically, we estimate

$$\begin{aligned} Y_{i,t} &= \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. Idiosyncratic volatility is the standard deviation of daily residuals from a Fama-French three factor regression in the year after index reconstitution (Ang et al., 2006). Short-term reversal beta is the level of the slope coefficient in a regression of daily excess returns on the short-term reversal factor. Bid-ask spread is the level of the average percentage bid-ask spread in the year after index reconstitution. Liquidity beta is from a regression of monthly excess returns on the market factor and the Pástor and Stambaugh liquidity factor. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variables are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Bid-ask spread	Liquidity beta	Short-term reversal beta	Idiosyncratic volatility
$\widehat{PO}$	0.90***	0.92***	0.69***	0.62***
	(4.56)	(3.67)	(3.69)	(2.79)
Bandwidth	300	300	300	300
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	17349	17868	17890	17890
R-squared (%)	2.07	0.37	0.42	6.03

Specifically, Haddad et al. (2021) show that aggregate demand curves for individual stocks become substantially less elastic when investors adopt passive investment strategies. Given this lower elasticity, we would expect to see more extreme price movements in individual stocks on average. However, measuring a stock's likelihood of extreme returns is inherently difficult since tail events are, by definition, rare. We thus rely on option-implied measures to quantify the effect of PO on a stock's tail risk.

We follow Bollerslev and Todorov (2011), who decompose the quadratic variation of the price process into a diffusive component and a jump component. These two risks have fundamentally different implications for market making capacity: diffusive risks can be hedged, whereas locally unpredictable jumps cannot. We estimate the perceived probability of rare large jump events using nonparametric risk-neutral (left and right) tail jump measures derived by Bollerslev and Todorov (2011), i.e., we compute the quantities

$$LT_t^{\mathbb{Q}}(k) \approx \frac{e^{r_{t,T}(T-t)}P_t(K)}{(T-t)F_t} \tag{2}$$

and

$$RT_t^{\mathbb{Q}}(k) \approx \frac{e^{r_{t,T}(T-t)}C_t(K)}{(T-t)F_t}, \tag{3}$$

where $C_t(K)$ ($P_t(K)$) is the price of a European call (put) option with 30 days to expiration and strike price K. F_t is the price of a forward contract expiring on the same date as the option, and, as in Bollerslev and Todorov (2011), $k \equiv \ln(K/F_t)$ is set to -0.1 and 0.1 for the left and right tail, respectively.

Intuitively, prices of short-maturity deep out-of-the-money options largely reflect the presence of perceived jumps, whereas the diffusive part of the price process plays a minor role. In practice, single name options are traded American-style, and thus observed option prices cannot be used to compute these measures directly. Therefore, we use the volatility surface to compute prices of equivalent European options, following Carr and Wu (2009) and Martin and Wagner (2019), among others. Appendix A.2 provides more details on the data and the computation of the tail jump measures.

To get started, Figure A.2 in Appendix A.2 plots the average of option-implied tail risk across stocks at the daily frequency from 1996-2021. There are two notable spikes, in both the right and left tail risk measure, one during the global financial crisis and one during the

Table IV: Impact of passive ETF ownership on tail risk measures. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on tail risk. Specifically, we estimate

$$\begin{aligned} Y_{i,t} &= \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $Y_{i,t}$ is given by the left and right option-implied tail jump risk measures following Bollerslev and Todorov (2011) (Equations (2) and (3)) and by the number of extreme daily return observations less than -10% or greater than 10%, respectively. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	$LT^{\mathbb{Q}}$	$RT^{\mathbb{Q}}$	$\#r_{d,i,t} < -10\%$	$\#r_{d,i,t} > 10\%$
$\widehat{PO}$	19.18***	19.40***	3.35***	2.28**
	(4.7)	(4.29)	(3.9)	(2.59)
Bandwidth	300	300	300	300
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	18620	18620	17368	17368
R-squared (%)	9.82	10.33	1.58	2.65

COVID-19 shock. Moreover, there seems to be a slight upward trend in both tail risks since 2008.

Based on the index experiment, we now regress average left and right tail risk in the twelve months after index reconstituion (i.e., from July to June) on instrumented passive ETF ownership. The first two columns of Table IV show that a one standard deviation increase in PO leads to a 19 percentage points higher implied tail risk for the left and right tails, confirming the hypothesis that stocks with high pasive ETF ownership are more prone to sharp price moves. This result suggests that shifts in PO are priced, at least in option markets.

Since option liquidity varies in the cross section of stocks, we now want to address the potential concern that our results on tail risk may be driven by outliers or low liquidity in options of some stocks. We thus also build an alternative measure of tail risk, which is based on simply counting the number of daily returns which exceed (fall below) a threshold of 10% (-10%). Specifically, we count the occurrence of such extreme returns between July (directly after the index reconstitution) and the end of June in the next year (before the next reconstitution) and use this count as our measure of tail risk. The third and fourth column in Table IV reports results for the second stage regression of these counts on instrumented PO. A one-standard deviation increase in passive ETF ownership significantly increases the occurrence of extreme daily returns by around 3 (i.e., we would observe three more of those extreme daily returns per year), which corroborates our findings on option-implied tail risk.

4. Why does passive ownership affect market quality?

We now test for the effect of PO on the second dimension of market quality, i.e. how variation in passive ownership affects price efficiency. In Section 4.1, we decompose stock return variation into a transitory component (noise) and permanent components (firm-specific news and market-wide news) based on a structural vector autoregression. We test how passive ownership affects the relative importance of these components and find that higher PO significantly *increases* the importance of (transitory) noise but *decreases* the importance of firm-specific information. To better understand this result, we further investigate if higher PO increases the sensitivity of stock returns to market-wide sentiment shocks (Section 4.2).

4.1. Passive ownership, news, and noise

Variance decomposition. We decompose stock return variance into different drivers based on the approach recently developed by Brogaard et al. (2022). Their variance decomposition is based on a structural vector autoregression (VAR) comprising stock returns, market returns, and order flow (based on signed trading volume as in Pástor and Stambaugh (2003)). A key idea in this setup is to differentiate between the permanent and transitory response of returns to shocks. Specifically, Brogaard et al. (2022) suggest to measure noise in returns as the transitory response of stock returns to shocks, whereas the long-term response of returns to shocks must be due to news about the market or firm-specific news.[12]

With noise being the transitory variation in returns, the long-run (permanent) response of

[12]For example, if the long-run response of a firm's stock return to market returns is, say, 1%, but there is a (transitory) overreaction so that the stock return first increases by 2% before settling on the long-run response of 1%, this would be interpreted as noise in this framework.

stock returns to innovations in market returns quantifies the importance of market-wide news. Likewise, the long-run responses of stock returns to innovations in returns and order flow provide a measure for the importance of firm-specific news. Note that Brogaard et al. (2022) further distinguish between private firm-level information (response to order flow shocks) and public firm-level information (response to firm-level returns).[13] For completeness, we also report these two components of firm-specific information in our results but are more agnostic about their nature and will mostly focus on firm-specific information as a whole.

We follow the estimation procedure outlined in Brogaard et al. (2022) and estimate a separate VAR for each stock in each year (using daily returns from July to June) and compute a variance decomposition for each stock-year to quantify the share of variance due to noise, firm-level news, and market-wide news based on the following return decomposition

$$r_t = \underbrace{a_0}_{\text{discount rate}} + \underbrace{\underbrace{\theta_{r_m}\varepsilon_{r_m,t}}_{\text{market info}} + \underbrace{\underbrace{\theta_x\varepsilon_{x,t}}_{\text{private info}} + \underbrace{\theta_r\varepsilon_{r,t}}_{\text{public info}}}_{\text{firm-specific info}}}_{\text{new information } (w_t)} + \underbrace{\Delta s_t}_{\text{noise}} \tag{4}$$

where we refer to Appendix A.3 for a precise explanation of the procedure and notation and to Brogaard et al. (2022) for computational details.[14] Across all stocks in the CRSP sample, market-wide information represents the smallest component, with an average share of 14.6%. Firm-specific information (public and private) and noise account for 65% and 20.5% of stock return variance, respectively.

Regression results. Turning to the effect of PO on the drivers of returns, we use the index experiment and perform second-stage regressions of the four variance shares detailed above on instrumented passive ETF ownership from the first stage.

Table V reports the results. To have a benchmark, we first report results for a regression of total stock return volatility on instrumented ETF ownership and find that a one-standard

[13]This distinction follows the microstructure literature (Hasbrouck, 1991a,b) and is based on the idea that order flow reveals private information by market participants. However, one can easily come up with a story in which order flow responds to public news as well (as also discussed in Brogaard et al. (2022)) so this distinction is unlikely to be exact.

[14]We replicate the main variance decomposition results in Brogaard et al. (2022) for our sample period and find results that are very similar to theirs.

Table V: Impact of passive ETF ownership on variance components. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on return variance component shares. Specifically, we estimate

$$\begin{aligned} VarianceShare_{i,t} &= \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $VarianceShare_{i,t}$ is represented by total variance and the variance shares obtained from the Brogaard et al. (2022) decomposition model, respectively. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics (in parentheses) are based on standard errors clustered at the firm level. Bandwidth indicates the number of firms at the bottom of the Russell 1000 and the top of the Russell 2000 included in our sample. Data are annual and cover index reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

		Variance shares			FirmInfo	
Dependent variable =	Total Variance	Noise	FirmInfo	MktInfo	PrivateInfo	PublicInfo
$\widehat{PO}$	12.79***	6.41***	-9.10***	2.69	-14.47***	5.37*
	(4.09)	(3.65)	(-2.84)	(0.94)	(-4.31)	(1.91)
Bandwidth	300	300	300	300	300	300
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17197	17197	17197	17197	17197	17197
R-squared (%)	3.02	0.34	1.76	2.60	0.20	2.12

deviation increase in ETF ownership causes 12.8 percentage points higher return variance, which essentially replicates the result of Ben-David et al. (2018). This effect is substantial given that average stock return variance amounts to 9.6% per year. In volatility terms, this increase would mean to go from an average volatility of around 31% to about 47% annually.

The variance decomposition now allows us to further understand this result. We report the effect of PO on the variance share of noise in the second column in Table V and find that the importance of noise rises significantly by 6.4 percentage points for a one standard deviation increase in passive ETF ownership. This number is high relative to the unconditional average of 15% attributable to noise. Higher passive ownership thus increases the share of transitory return variance.

The next column in Table V reports second-stage regression results for the share of firm specific information (FirmInfo), and we find that higher PO *decreases* the importance of this kind of news significantly by about nine percentage points for a one standard deviation increase in PO. This effect is economically large, given that firm-specific information on average accounts for about 63% of return variance.[15]

The fourth column in Table V (MktInfo) shows that there is a small, positive effect of PO on the importance of market-wide information, but this effect is statistically insignificant. Importantly, this result relates to the *permanent* response of stock returns to market news and does not necessarily mean that stocks do not react more strongly to market news in a transitory fashion. In fact, Figure 3 shows that higher PO increases the transitory response of stock returns to market news for a few days, before the effect levels out and converges to the long-run response. This increase in the transitory response to market news contributes to the higher share of noise documented above.

Finally, the last two columns report results for the decomposition of firm-specific news into private and public news. As mentioned above, this decomposition is not straightforward to interpret but, based on the interpretation in Brogaard et al. (2022), one would conclude that the decrease in firm-specific news is almost completely driven by a decrease in *private* firm-specific information (-14.5 percentage points with a t-statistic of -4.3) whereas the importance of *public* firm-specific news is slightly positive but only marginally significant ($t = 1.9$).[16]

Taken together, these results may help rationalize our findings in Section 3. The higher share of noise in returns brought about by a larger amount of noise per se reduces market making capacity and leads to more reversal, especially for liquidity-driven stock price changes. The lower share of firm-specific information suggests a reduction of the informa-

[15] One might argue that finding a lower share of firm-specific information is mechanical due to the fact that noise increases (i.e. a rise in the denominator in the variance decomposition). However, from an empirical point of view, this relation does not necessarily mean that each individual type of news has to decline. In fact, as Table V shows, the importance of public information and market-wide information increases instead of declines. Moreover, from a conceptual point of view, higher PO might also lead to more informative prices, e.g. due to the ability to circumvent short-sale constraints (Huang, O'Hara, and Zhong, 2021). Our results suggest that such effects are not strong enough to counteract the effect of higher noise, indicating an overall decrease in price informativeness and that this effect largely stems from the lower importance of private firm-specific news.

[16] This marginal increase in the importance of public news might explain why recent research has documented that more ETF ownership is associated with a stronger and/or faster response to earnings announcements, see, e.g., Glosten et al. (2021) or reduced profits to anomalies based on public information (Filippou et al., 2022).

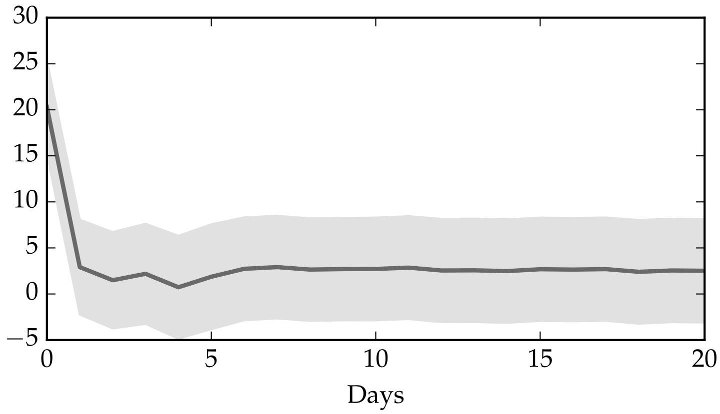


Figure 3: Impact of instrumented passive ETF ownership on market information shares. This figure shows slope coefficients from regressions of passive ETF ownership on a stock's market information shares. Specifically, we estimate

$$\begin{aligned} MktInfoShare^{j}_{i,t} &= \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is instrumented passive ETF ownership (Appel et al., 2019). $MktInfoShare^{j}_{i,t}$ is estimated using the structural cumulative impulse response of the stock return to a unit shock in market returns for day $j = 1, 2, \ldots, 19, 20$ using the variance decomposition from Brogaard et al. (2022). Data are annual and cover index reconstitutions from 2000 to 2020.

tional efficiency of prices and is in line with earlier theoretical work (e.g. Baruch and Zhang, 2022; Bhattacharya and O'Hara, 2018; Vayanos and Whoolley, 2023).[17] A common theme is that the effect of PO on market efficiency can vary between the micro (firm-specific news) and the macro level (market-wide news). Our results speak to this theoretical literature in that it suggests that stock prices become significantly less efficient at the micro level but that they do not necessarily become less efficient in impounding information into prices at the macro level since we also find that the importance of market-wide news is not adversely affected by an increase in PO.

4.2. *Sentiment and mispricing*

We now aim for a better understanding of how and why passive ETF ownership increases transitory noise in returns. A potential explanation for our findings in Section 4.1 could be

[17] This result is also generally in line with the notion that stocks locked up by passive investors cannot be used to trade on firm-specific information anymore (Israeli, Lee, and Sridharan, 2017) which might lower price efficiency.

Table VI: Impact of passive ETF ownership on stock return sensitivity to changes in sentiment and mispricing. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on sentiment betas and average mispricing scores. Specifically, we estimate

$$\begin{aligned} Y_{i,t} \quad = \quad & \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ & + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. The dependent variable is a stock's beta with respect to changes in an aggregate sentiment measure or its mispricing score according to Stambaugh et al. (2012). Measures of aggregate sentiment are the American Association of Individual Investors (AAII) sentiment survey, the Baker and Wurgler (2007) sentiment index, and UMC consumer sentiment, respectively. AAII sentiment is a monthly average of the bull-bear-spread, defined as the percentage of bullish investors minus the percentage of bearish investors. Sentiment betas are from a regression of monthly excess returns on orthogonalized changes in sentiment and the market factor. Orthogonalization is performed with respect to the Fama-French three factor model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	AAII sentiment beta	BW sentiment beta	UMC sentiment beta	Mispricing score
$\widehat{PO}$	0.46**	0.42**	0.54**	0.56**
	(2.04)	(2.03)	(2.04)	(2.09)
Bandwidth	300	300	300	300
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	17868	17868	17868	12347
R-squared (%)	0.29	0.26	0.19	0.90

that passive ETFs attract more unsophisticated (noise) traders that trade on short-term, non-fundamental information (see Ben-David et al., 2018), because of, e.g., the ease of trading and the low costs associated with passive ETFs. If so, we would expect that stocks with more passive ETF ownership are more exposed to measures of transitory shocks and are more mispriced. We now test these hypotheses by estimating the impact of PO on a stock's exposure to (aggregate) sentiment fluctuations and its mispricing.

We employ three commonly used measures of sentiment, the American Association of Individual Investors Investor (AAII) sentiment survey, the Baker and Wurgler (2007) sentiment index, and University of Michigan (UMC) consumer sentiment (see, e.g. Greenwood and Shleifer (2014)).[18] The AAII survey reports the weekly percentages of investors who have bullish, neutral, or bearish views on the stock market for the next six months. Following Greenwood and Shleifer (2014), we use the monthly average difference between the shares of bullish and bearish investors as the measure for AAII sentiment.

Separately for each stock and twelve month period, we regress excess returns on changes in sentiment and the market factor, i.e., we estimate

$$r_{i,t}^{e} = \alpha_i + \beta_i^{SENT} \Delta SENT_t + \beta_i^{Mkt} r_t^{Mkt-RF} + \varepsilon_{i,t}, \tag{5}$$

where β_i^{SENT} represents the exposure to changes in aggregate sentiment.[19] We then use the estimated β_i^{SENT} as a dependent variable in a regression on instrumented passive ETF ownership and report results in the first three columns of Table VI. The results are clear-cut and show that higher PO significantly increases stock return sensitivity to changes in market-wide sentiment with t-statistics of slightly above 2 for all three sentiment measures.

To test whether changes in PO affect the mispricing of stocks, we further run regressions of a stock's mispricing score based on Stambaugh et al. (2012) on instrumented passive ETF ownership and a number of control variables. This mispricing score for individual stocks is the average normalized rank across eleven different anomaly variables.[20] Stocks with higher score are more "overpriced". The fourth column in Table VI reports results for this exercise, and we find a significantly positive effect of PO on mispricing, i.e., an increase in passive ETF ownership leads, on average, to more pronounced overpricing.

We next consider an alternative mechanism suggested in the recent literature. Specifically, Ben-David et al. (2023) show that ETF providers respond to investor sentiment by catering to demand for popular and overvalued investment themes. This could be an important driver of our results on sentiment betas and mispricing, as stocks that attract investors with overoptimistic beliefs tend to be included in newly launched specialized ETFs. To disentangle this effect from our hypothesis that passive ETFs attract noise traders, we perform

[18] AAII sentiment survey data are taken from `www.aaii.com/sentimentsurvey/sent_results`, Baker and Wurgler sentiment is obtained from Jeffrey Wurgler's website (`https://pages.stern.nyu.edu/~jwurgler`), and UMC sentiment is available at `https://fred.stlouisfed.org/series/UMCSENT`.

[19] We first orthogonalize changes in sentiment with respect to the Fama-French three factor model using a full-sample time-series regression. Our results remain significant using raw changes in sentiment instead.

[20] We obtain mispricing scores from Robert F. Stambaugh's website (`https://finance.wharton.upenn.edu/~stambaug`).

an additional robustness check in which we exclude ETFs launched within the last five years from the sample. This way, recent "hot investment themes" packaged into ETFs are unlikely to affect our results. Figure IA.1 shows the evolution of the cross-sectional average PO within a quarter while restricting the sample to funds that started reporting their holdings more than five years ago. Excluding these funds reduces PO by 0.5 percentage points on average, but maintains a similar overall trend over time. Finally, we regress instrumented ownership of passive ETFs, that are more than five years old, on sentiment betas and mispricing. Table IA.5 shows that the size of the effect as well as the statistical significance remain virtually unchanged. We take this as evidence that our main findings cannot be explained by the inclusion of popular stocks, possibly mispriced stocks, into recently launched ETFs.

5. Robustness

This section provides additional results and robustness checks, for which most tables are outsourced to a separate internet appendix. We examine the importance of ownership by passive mutual funds as opposed to ETFs (Section 5.1) and provide evidence for results based on the Russel index experiment using different samples as suggested by Appel, Gormley, and Keim (2020) (Section 5.2).

5.1. *Including passive mutual fund ownership*

Our results so far are based on passive *ETF* ownership, since passive ETFs volumes have seen the largest growth rates in recent years. Moreover, passive ETFs offer liquidity in a way that mutual funds do not (e.g., by being exchange-traded on a more or less continuous basis, not just once at the end of the day), which makes them attractive for short-term trading and non-fundamental speculation (e.g. Ben-David et al., 2018). A key question thus is whether our key results are confined to passive ETF ownership or whether other passive vehicles may have similar effects (see Pavlova and Sikorskaya, 2023, for a comprehensive overview of passive investing). A thorough analysis of this question is beyond the scope of our paper but we report key results from our analysis for an extended measure of passive ownership here, covering both passive mutual funds and passive ETFs.[21]

[21] Following Appel et al. (2016), we identify passive funds based on their fund names or by using the *index_fund_flag* from the CRSP Mutual Fund Database.

We summarize key results in Tables IA.6, IA.7, and IA.8 in the internet appendix. We find that assignment of a firm to the Russell 2000 increases passive ownership including passive mutual funds (Table IA.6), that the effect of passive ownership on the importance of noise and firm-specific information from the variance decomposition, if anything, increases (IA.7), and that passive ownership significantly increases return reversals (IA.8). Taken together, these results imply that the findings documented in our main analyses are not really specific to passive ETFs, but seem to hold for other forms of passive investing as well.

5.2. Alternative samples

In this section we show that our results using the Russell index experiment are not sensitive to the choice of the sample. While most results reported in the main text focus on bandwidths of 300 firms, we present tables with varying bandwidths in the Internet Appendix (Tables IA.9 to IA.23). Almost all results remain similar in size and statistical significance when considering bandwidths between 200 and 500. Only in some cases statistical significance weakens for small bandwidths of 200 firms, or large bandwidths of 500 firms.

Sample selection also depends on whether stocks are ranked based on total or float-adjusted market capitalization (Appel et al., 2020). The former is used by Russell to assign stocks to each index, whereas the latter determines within-index weights. The advantage of using float-adjusted market capitalization ranks is that stocks with lowest (highest) weights in the Russell 1000 (2000) would be selected for our analyses. However, float-adjusted market capitalization is highly correlated with liquidity and inside-ownership (Appel et al., 2020). We therefore decided to rank stocks based on total market capitalization, but now also show that results are similar using within-index ranks. Table IA.24 shows the first stage regression of ETF ownership on assignment to the Russell 2000. Coefficients are slightly higher, and statistical significance increases. Table IA.25 shows our regression of variance shares on instrumented ETF ownership. Our key findings still hold, albeit slightly lower in magnitude and statistical significance.

6. Conclusion

Our study contributes to the understanding of the impact of passive ETF ownership on stock return drivers and market efficiency. We find that higher passive ETF ownership significantly decreases the informativeness of stock prices by increasing the importance of non-fundamental return noise and reducing the contribution of firm-specific information. This highlights the potential implications for market efficiency and the role of active investors in processing fundamental information.

Our findings have important implications for policymakers. The rise of passive investing and the associated increase in passive ETF ownership may warrant a closer examination of its effects on market dynamics. Policymakers should consider the potential trade-offs between the benefits of lower transaction costs and the potential costs of reduced price efficiency and market-making capacity. Balancing the growth of passive investing with maintaining informativeness of asset prices could turn out to become a crucial challenge for regulators and market participants.

Furthermore, our study opens avenues for future research. First, exploring the channels through which passive ETF ownership affects return noise and firm-specific information can shed light on the underlying economic mechanisms. Investigating the impact of different passive investment strategies, such as factor-based or thematic ETFs, on market efficiency and price informativeness would provide a more nuanced understanding of the impact of passive ownership in a more and more diverse passive investing landscape.

Our results on the impact of passive ownership on liquidity, exposure to sentiment shocks, and tail risk also highlight the need for more research on the implications of passive ETF ownership for asset pricing models and risk management. While our results using the index experiment identify the causal impact of changes in passive ownership on comparable stocks close to the index cut-off, more work is needed to understand the "global" impact of passive ownership on market quality. Likewise, understanding how increased share of noise and a reduced share of firm-specific information affect asset pricing factors and risk premiums could enhance our understanding of the link between passive investing and asset valuation.

References

Agarwal, V., P. Hanouna, R. Moussawi, and C. W. Stahel (2018). Do ETFs increase the commonality in liquidity of underlying stocks? In *28th Annual Conference on Financial Economics and Accounting, Fifth Annual Conference on Financial Market Regulation*.

Ahn, B. H. and P. N. Patatoukas (2022). Identifying the effect of stock indexing: Impetus or impediment to arbitrage and price discovery? *Journal of Financial and Quantitative Analysis 57*(5), 2022–2062.

Ang, A., R. J. Hodrick, Y. Xing, and X. Zhang (2006). The cross-section of volatility and expected returns. *Journal of Finance 61*(1), 259–299.

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

Appel, I. R., T. A. Gormley, and D. B. Keim (2016). Passive investors, not passive owners. *Journal of Financial Economics 121*(1), 111–141.

Appel, I. R., T. A. Gormley, and D. B. Keim (2019). Standing on the shoulders of giants: The effect of passive investors on activism. *Review of Financial Studies 32*(7), 2720–2774.

Azar, J., M. Duro, I. Kadach, and G. Ormazabal (2021). The big three and corporate carbon emissions around the world. *Journal of Financial Economics 142*(2), 674–696.

Baker, M. and J. Wurgler (2007). Investor sentiment in the stock market. *Journal of Economic Perspectives 21*(2), 129–151.

Ball, R. and P. Brown (1968). An empirical evaluation of accounting income numbers. *Journal of Accounting Research 6*, 159–178.

Ball, R., J. Gerakos, J. T. Linnainmaa, and V. Nikolaev (2016). Accruals, cash flows, and operating profitability in the cross section of stock returns. *Journal of Financial Economics 121*(1), 28–45.

Banz, R. W. (1981). The relationship between return and market value of common stocks. *Journal of Financial Economics 9*(1), 3–18.

Baruch, S. and X. Zhang (2022). The distortion in prices due to passive investing. *Management Science 68*, 6219–6234.

Ben-David, I., F. Franzoni, B. Kim, and R. Moussawi (2023). Competition for attention in the etf space. *Review of Financial Studies 36*(3), 987–1042.

Ben-David, I., F. Franzoni, and R. Moussawi (2018). Do ETFs increase volatility? *Journal of Finance 73*(6), 2471–2535.

Ben-David, I., F. Franzoni, and R. Moussawi (2019). An improved method to predict assignment of stocks into russell indexes. Technical report, National Bureau of Economic Research.

Bhattacharya, A. and M. O'Hara (2018). Can ETFs increase market fragility? Effect of information linkages in ETF markets. Working Paper.

Bollerslev, T. and V. Todorov (2011). Tails, fears, and risk premia. *Journal of Finance 66*(6), 2165–2211.

Bond, P. and D. Garcia (2022). The equilibrium consequences of indexing. *Review of Financial Studies 35*(7), 3175–3230.

Brogaard, J., T. H. Nguyen, T. J. Putnins, and E. Wu (2022). What moves stock prices? The roles of news, noise, and information. *Review of Financial Studies 35*(9), 4341–4386.

Brown, D. C., S. W. Davies, and M. C. Ringgenberg (2021). ETF arbitrage, non-fundamental demand, and return predictability. *Review of Finance 25*(4), 937–972.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market liquidity and funding liquidity. *Review of Financial Studies 22*(6), 2201–2238.

Carr, P. and L. Wu (2009). Variance risk premiums. *Review of Financial Studies 22*(3), 1311–1341.

Chang, Y.-C., H. Hong, and I. Liskovich (2015). Regression discontinuity and the price effects of stock market indexing. *Review of Financial Studies 28*(1), 212–246.

Chen, A. Y. and T. Zimmermann (2022). Open source cross-sectional asset pricing. *Critical Finance Review 27*(2), 207–264.

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

Cooper, M. J., H. Gulen, and M. J. Schill (2008). Asset growth and the cross-section of stock returns. *Journal of Finance 63*(4), 1609–1651.

Dai, W., M. Medhat, R. Novy-Marx, and S. Rizova (2024). Reversals and the returns to liquidity provision. *Financial Analysts Journal*, 1–30.

Fama, E. F. and K. R. French (1992). The cross-section of expected stock returns. *Journal of Finance 47*(2), 427–465.

Fama, E. F. and J. D. MacBeth (1973). Risk, return, and equilibrium: Empirical tests. *Journal of Political Economy 81*, 607–636.

Filippou, I., S. He, and G. Zhou (2022). ETFs, anomalies and market efficiency. Working Paper.

Gao, X. and J. R. Ritter (2010). The marketing of seasoned equity offerings. *Journal of Financial Economics 97*(1), 33–52.

Gârleanu, N. and L. H. Pedersen (2022, 06). Active and Passive Investing: Understanding Samuelson's Dictum. *The Review of Asset Pricing Studies 12*(2), 389–446.

Glosten, L., S. Nallareddy, and Y. Zou (2021). ETF activity and informational efficiency of underlying securities. *Management Science 67*(1), 22–47.

Glosten, L. R. and P. R. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of Financial Economics 14*(1), 71–100.

Greenwood, R. and A. Shleifer (2014). Expectations of returns and expected returns. *Review of Financial Studies 27*(3), 714–746.

Grossman, S. J. and M. H. Miller (1988). Liquidity and market structure. *The Journal of Finance 43*(3), 617–633.

Haddad, V., P. Huebner, and E. Loualiche (2021). How competitive is the stock market? Theory, evidence from portfolios, and implications for the rise of passive investing. In *Proceedings of Paris December 2021 Finance Meeting EUROFIDAI-ESSEC*.

Hasbrouck, J. (1991a). Measuring the information content of stock trades. *Journal of Finance 46*(1), 179–207.

Hasbrouck, J. (1991b). The summary informativeness of stock trades: An econometric analysis. *Review of Financial Studies 4*(3), 571–595.

Ho, T. and H. R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics 9*(1), 47–73.

Hou, K., C. Xue, and L. Zhang (2020). Replicating anomalies. *Review of Financial Studies 33*(5), 2019–2133.

Huang, S., M. O'Hara, and Z. Zhong (2021). Innovation and informed trading: Evidence from industry ETFs. *Review of Financial Studies 34*(3), 1280–1316.

Israeli, D., C. Lee, and S. A. Sridharan (2017). Is there a dark side to exchange traded funds? An information perspective. *Review of Accounting Studies 22*(3), 1048–1083.

Jappelli, R. (2023). Dynamic asset pricing with passive investing. Working Paper, SAFE Frankfurt.

Jegadeesh, N. (1990). Evidence of predictable behavior of security returns. *Journal of Finance 45*(3), 881–898.

Jegadeesh, N. and S. Titman (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance 48*(1), 65–91.

Jiang, G. J. and Y. S. Tian (2005). The model-free implied volatility and its information content. *Review of Financial Studies 18*(4), 1305–1342.

Jiang, H., D. Vayanos, and L. Zheng (2020). Passive investing and the rise of mega-firms. Technical report, National Bureau of Economic Research.

Kyle, A. S. (1985). Continuous auctions and insider trading. *Econometrica 53*, 1315–1335.

Martin, I. W. and C. Wagner (2019). What is the expected return on a stock? *Journal of Finance 74*(4), 1887–1929.

Medhat, M. and M. Schmeling (2022). Short-term momentum. *Review of Financial Studies 35*(3), 1480–1526.

Moskowitz, T. J. and M. Grinblatt (1999). Do industries explain momentum? *Journal of Finance 54*(4), 1249–1290.

Newey, W. K. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica 55*, 703–708.

O'Hara, M. and M. Ye (2011). Is market fragmentation harming market quality? *Journal of Financial Economics 100*(3), 459–474.

Pástor, L. and R. F. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*(3), 642–685.

Pavlova, A. and T. Sikorskaya (2023). Benchmarking intensity. *Review of Financial Studies 36*(3), 859–903.

Peress, J. and D. Schmidt (2020). Glued to the tv: Distracted noise traders and stock market liquidity. *The Journal of Finance 75*(2), 1083–1133.

Schmidt, C. and R. Fahlenbrach (2017). Do exogenous changes in passive institutional ownership affect corporate governance and firm value? *Journal of Financial Economics 124*(2), 285–306.

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

Vayanos, D. and P. Whoolley (2023). Asset management as creator of market inefficiency. *Atlantic Economic Journal 51*, 1–11.

Appendix

A.1. Identification strategy: Index reconstitutions

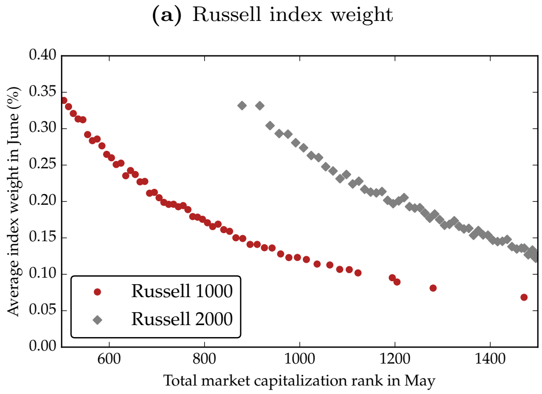
Our identification strategy relies on annual reconstitutions of the Russell stock indices.[22] At the end of each June, Russell uses end of May total market capitalization to assign the largest 1,000 eligible stocks to the Russell 1000 index, and the next 2,000 stocks to the Russell 2000 index. Within each index, constituents are weighted by their end of June float-adjusted market capitalization. As a result, firms that appear on top of the Russell 2000 instead of the bottom of the Russell 1000 receive significantly higher index weights and consequently significantly higher ETF ownership. Importantly, this discontinuity does not apply to other stock characteristics, as these firms are relatively similar in terms of their size and liquidity. To limit the amount of stocks switching indexes each year, Russell has updated their index methodology in 2006.[23] Since then, stocks are not allowed to switch indexes if the stock had already changed index in the previous year or if its market capitalization falls within a certain range around the cutoff ("banding" rule). The discontinuity in index weights between Russell 1000 and Russell 2000 stocks and the resulting differences in ETF ownership are visualized in Figure A.1, akin to Pavlova and Sikorskaya (2023). While Panel (b) of Figure A.1 shows a pronounced discontinuity in passive ETF ownership, we note that this is not in the form of a sudden jump exactly at the index cutoff (Rank of 1000) due to the "banding rule."

In our empirical analysis, we repeatedly carry out two-stage least squares instrumental variable (IV) estimations to quantify the causal effect of passive ETF ownership on the outcome variable of interest. In our implementation, we follow the approach outlined by Appel et al. (2019). They argue that assignment to the Russell 2000 index is a random event for firms close to the cutoff after conditioning on the determinants of index membership.[24] The first stage of our two-stage IV approach estimates the effect of assignment to the Russell 2000 on passive ETF ownership and is

[22] See, e.g. Appel et al. (2016); Schmidt and Fahlenbrach (2017); Appel et al. (2019); Ben-David et al. (2018, 2019); Pavlova and Sikorskaya (2023) for other work based on this setting.

[23] See Section 6.10 in `https://research.ftserussell.com/products/downloads/Russell-US-indexes.PDF`

[24] Russell assigns stocks to each index based on a proprietary measure of end of May total market capitalization, which is unobservable. However, it is crucial to robustly control for this variable to avoid an omitted variable bias (Appel et al., 2020). We therefore rely on an alternative variable constructed by Ben-David et al. (2019), which has a 99.7% success rate in predicting index assignment.



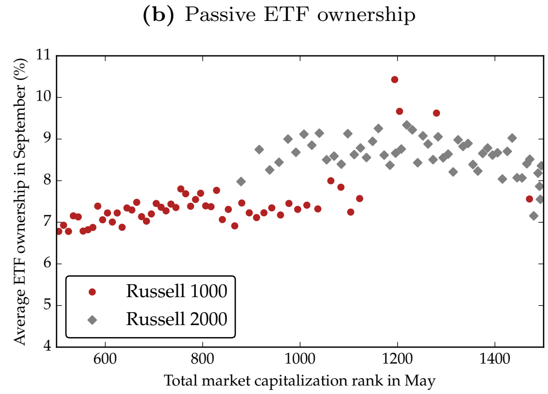


Figure A.1: Discontinuities around the Russell 1000/2000 cutoff. The figure shows average index weights and passive ETF ownership for firms ranked between 500 and 1500 in terms of end of May total market capitalization. We compute total market capitalization using the method of Ben-David et al. (2019). ETF ownership is the percentage share of a stock's market capitalization held by passive ETFs at the end of September. Index weights are based on float-adjusted market capitalization at the end of June. Averages are taken across bins of 20 stocks. The sample is annual and covers all firms included in the Russell 3000(E) from 2000 to 2020.

given by

$$\begin{aligned} PO_{i,t} \quad = \quad & \eta + \lambda R2000_{i,t} + \sum_{n=1}^{N} \chi_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ & + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $R2000_{i,t}$ is an indicator for a stock's membership in the Russell 2000 in July in year t. $Mktcap_{i,t}$ controls for end-of-May total market capitalization and $Float_{i,t}$ for end-of-June float-adjusted market capitalization used by Russell to set index weights.[25] $Band_{i,t}$ is an indicator for whether a firm's end-of-May market capitalization is within the banding range.[26] $R2000_{i,t-1}$ is one if a stock was assigned to the Russell 2000 in the previous year and zero otherwise. In our main analysis, we select 300 stocks on each side of each cutoff based on end-of-May total market capitalization ranks. In the internet appendix we show that our results remain robust across a wider range of bandwidths. Furthermore, we refrain from using ranks based on Russell's float-adjusted market capitalization to select our sample, since this quantity is substantially correlated with stock liquidity (Appel et al., 2020). In Section 5 of the main paper we show that our main results are not sensitive to this choice.

[25]Russell determines index weights using float-adjusted market capitalization to reduce trading costs of index trackers. It is important to also control for this second measure of size, since stocks with low liquidity and high-inside ownership will be given lower index weights (Appel et al., 2020).

[26]See Internet Appendix OC of Azar, Duro, Kadach, and Ormazabal (2021) for more details on the construction of this variable.

Table A.1: Impact of index assignment on passive ETF ownership. This table presents the first stage of an instrumental variable regression of ETF ownership on membership in the Russell 2000 index and further controls. Specifically, we estimate

$$\begin{aligned} PO_{i,t} \quad &= \quad \eta + \lambda R2000_{i,t} + \sum_{n=1}^{N} \chi_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $PO_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t. $R2000_{i,t}$ is our instrument and indicates a stock's assignment to the Russell 2000 at the end of June in year t. $Mktcap_{i,t}$ controls for end of May market capitalization computed using the method of Ben-David et al. (2019) and $Float_{i,t}$ is float-adjusted market capitalization in June used by Russell to determine index weights. To account for Russell's banding rule Appel et al. (2019) add three further controls: an indicator for a stock's membership in the Russell 2000 in the previous year ($R2000_{i,t-1}$), an indicator for whether a stock falls within a certain range close to the Russell cutoff ($Band_{i,t}$), and their interaction. The dependent variable is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Passive ETF ownership			
	(1)	(2)	(3)	(4)
R2000	0.13***	0.14***	0.15***	0.16***
	(6.54)	(7.25)	(7.87)	(8.81)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	14787	19707	24107	28506
R-squared (%)	0.45	0.42	0.38	0.45

Table A.1 presents results of the first stage regression. To ease interpretation, we follow Ben-David et al. (2018) and Appel et al. (2019) and divide the dependent variable by its standard deviation, measured across the pooled sample of stocks within the bandwidth. Consequently, second stage point estimates reflect the change in the response variable caused by a one standard deviation increase in passive ETF ownership. Furthermore, we include year fixed effects to control for the general increase in ETF ownership over time (see Figure 2 in the main text). The first stage shows that the effect of assignment to the Russell 2000 index ($R2000$) on passive ETF ownership ranges between 12.5% to 12.8% of one standard deviation, which roughly translates into an increase by 1 percentage point. The effect is statistically significant across all bandwidths, which confirms the validity of our instrument.

In the second stage, we regress an outcome variable of interest, which is measured during the year *after* index reconstitution, on the fitted value of ETF ownership from the first stage. The corresponding regression is

$$\begin{aligned} Y_{i,t} &= \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}. \end{aligned}$$

The outcome variable Y differs across applications, and we detail the construction of this outcome variable separately for each application in the main text.

A.2. Computing option-implied tail risk

This section provides details on the construction of our option-based tail risk measures. We use data on options with 30 days to expiration from the OptionMetrics volatility surface file (`vsurfd`). Implied volatilities are calculated by OptionMetrics using a binomial tree model which accomodates the early exercise premium.[27] Closing prices of underlyings are from the security price file (`secprd`). The risk-free interest rate is obtained from the zero curve file (`zerocd`). Continuous dividend yields over the previous 252 trading days are calculated as the difference between with and without dividend log returns from CRSP. Finally, data from OptionMetrics is matched to CRSP using the the linking file (`opcrsphist`) provided by WRDS.

We apply several filters to the options data. First, we remove all observations with zero or missing strike price. Second, we remove each stock-day for which there are less than five implied volatility observations available. Last, we eliminate all in-the-money options (such that we focus on $\Delta^{\text{put}} > -0.5$ and $\Delta^{\text{call}} \leq 0.5$).

The steps to compute option-implied tail jump risk are as follows.

1. Compute historical dividend yields and replace missing observations by 0.
2. Linearly interpolate the zero curve to obtain the risk-free rate matching each option expiration.
3. Calculate theoretical forward prices $F_t = S_t e^{(r_{t,T} - y_t)(T-t)}$, where $(T-t) = 30/365$.
4. Following Jiang and Tian (2005) and Carr and Wu (2009), combine out-of-the-money call and put implied volatilities at-the-money and interpolate them using a cubic spline across log-moneyness levels. Log-moneyness is defined as $k = \ln(K/F)$.

[27] See IvyDB File and Data Reference Manual, Version 5.0

5. Employ constant extrapolation of implied volatilities beyond the observable strike range using observations from the lowest and highest available log-moneyness.

6. Compute European call (put) prices with log-moneyness of $k = 0.1$ ($k = -0.1$), using the Black-Scholes-Merton formula and taking into account the continuous dividend yield.

The Bollerslev and Todorov (2011) tail risk measures are then computed as $LT_t^{\mathbb{Q}} = \frac{e^{r_{t,T}(T-t)}P_t(K)}{(T-t)F_t}$ and $RT_t^{\mathbb{Q}} = \frac{e^{r_{t,T}(T-t)}C_t(K)}{(T-t)F_t}$.

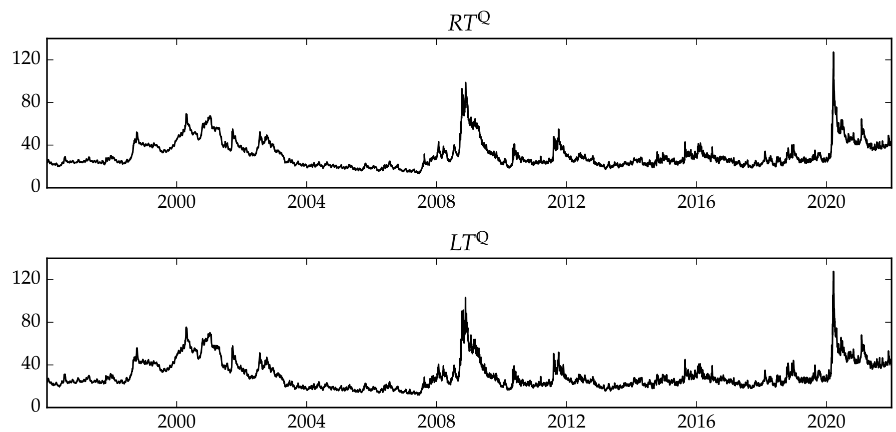


Figure A.2: Implied jump tail measures. This figure shows the (cross-sectional) average option-implied tail risk. Data are at the daily frequency and cover all underlyings included in the volatility surface file from 1996 to 2021.

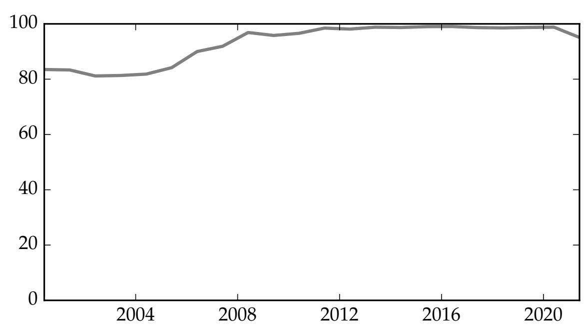


Figure A.3: Options data coverage around the Russell cutoffs. This figure shows the percentage of firms in our sample for which options data is available. The sample spans 300 firms on each side of each Russell cutoff. Data are annual from June 1997 to December 2021.

A.3. Variance decompositions

As suggested by Brogaard et al. (2022), we estimate a fifth-order structural VAR model[28]

$$
\begin{aligned}
r_{m,t} &= a_0 + \sum_{l=1}^{5} a_{1,l} r_{m,t-l} + \sum_{l=1}^{5} a_{2,l} x_{t-l} + \sum_{l=1}^{5} a_{3,l} r_{t-l} + \varepsilon_{r_m,t} \\
x_t &= b_0 + \sum_{l=0}^{5} b_{1,l} r_{m,t-l} + \sum_{l=1}^{5} b_{2,l} x_{t-l} + \sum_{l=1}^{5} b_{3,l} r_{t-l} + \varepsilon_{x,t} \\
r_t &= c_0 + \sum_{l=0}^{5} c_{1,l} r_{m,t-l} + \sum_{l=0}^{5} c_{2,l} x_{t-l} + \sum_{l=1}^{5} c_{3,l} r_{t-l} + \varepsilon_{r,t}
\end{aligned}
\tag{A.1}
$$

where $r_{m,t}$ is the daily market return, x_t is firm-level order flow, i.e. the signed dollar volume of a stock (Pástor and Stambaugh, 2003), and r_t denotes firm-level stock returns.[29] Unexpected innovations (shocks) are given by $\varepsilon_{r_m,t}, \varepsilon_{x,t},$ and $\varepsilon_{r,t}$. Furthermore, the moving-average representation of the structural VAR provides us with permanent cumulative stock return responses to unit shocks, which we denote by $\theta_{r_m}, \theta_x,$ and θ_r. As in Brogaard et al. (2022), we use the cumulative return response 15 days after each shock to obtain a measure for long-run responses.[30] The estimated parameters can be put together as follows

$$
r_t = \underbrace{a_0}_{\text{discount rate}} + \underbrace{\underbrace{\theta_{r_m}\varepsilon_{r_m,t}}_{\text{market info}} + \underbrace{\underbrace{\theta_x \varepsilon_{x,t}}_{\text{private info}} + \underbrace{\theta_r \varepsilon_{r,t}}_{\text{public info}}}_{\text{firm-specific info}}}_{\text{new information } (w_t)} + \underbrace{\Delta s_t}_{\text{noise}}
\tag{A.2}
$$

where Δs_t is noise, the net transitory return which is left after subtracting the permanent components (a_0 and w_t). Based on this return representation, we compute variance shares as normalized variance components for the share due to market information, firm-specific information, and the share due to noise:

[28] Specifically, we first estimate a reduced-form VAR and then recover structural parameters using reduced-form error covariances. The Internet Appendix of Brogaard et al. (2022) provides detailed steps on how to obtain the structural innovations, impulse-response functions and further details on the estimation that we follow in our estimation.

[29] Signed dollar volume as a proxy for order flow is computed as in Pástor and Stambaugh (2003). We adjust the trading volume of NASDAQ stocks prior to 2004 following Gao and Ritter (2010).

[30] Table IA.26 in the Internet Appendix shows that our results are virtually unchanged using the response after 60 days instead of 15 days.

$$
\begin{aligned}
\text{Market Information Share} &= \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2 / \left(\sigma_w^2 + \sigma_s^2\right) \\
\text{Private Information Share} &= \theta_x^2 \sigma_{\varepsilon_x}^2 / \left(\sigma_w^2 + \sigma_s^2\right) \\
\text{Public Information Share} &= \theta_r^2 \sigma_{\varepsilon_r}^2 / \left(\sigma_w^2 + \sigma_s^2\right) \\
\text{Noise Share} &= \sigma_s^2 / \left(\sigma_w^2 + \sigma_s^2\right).
\end{aligned} \tag{A.3}
$$

where the share of firm-specific news reported in Table V is given by the sum of the private and public information share, i.e., $(\theta_x^2 \sigma_{\varepsilon_x}^2 + \theta_r^2 \sigma_{\varepsilon_r}^2) / \left(\sigma_w^2 + \sigma_s^2\right)$.

As in Brogaard et al. (2022), we only retain observations with non-missing return, volume, and price date, require at least 20 days of observations per stock-year and winsorize variance components at the 5% and 95% level each year.

Internet Appendix for

Passive Investing and Market Quality

(not for publication)

IA.1. Tables and figures omitted from the main text

Table IA.1: Variable definitions.

Variable	Description	Source
AAII sentiment beta	We measure AAII sentiment as monthly average difference between the percentage of bullish and bearish investors. Next, we obtain orthogonalized sentiment as the residual from a full-sample regression of changes in sentiment on the Fama-French three factor model. Sentiment betas are from regressions of monthly excess stock returns on orthogonalized sentiment and the market factor, conducted separately for each stock and twelve month period.	CRSP, AAII Website
Asset growth (Cooper, Gulen, and Schill, 2008)	Asset growth is the percentage change in total assets (AT).	Open Asset Pricing (CZ, 2022)
Baker and Wurgler sentiment beta	We regress monthly excess stock returns on changes in Baker and Wurgler sentiment (orthogonalized with respect to the Fama-French three factor model) and the market factor. We require at least 10 non-missing observations.	CRSP, Jeffrey Wurgler's Website
Beta (Fama and MacBeth, 1973)	Market beta is estimated by a rolling regression of monthly excess returns on the excess return of the CRSP value-weighted market portfolio. We use a window size of 60 months and require at least 20 non-missing observations.	Open Asset Pricing (CZ, 2022)
Bid-ask spread	We measure bid-ask spread as the difference between the ask price and the bid price, divided by the mid price.	CRSP
CAR	Cumulative announcement return is the sum of daily abnormal returns from one day prior to one day after an earnings announcement. We compute abnormal returns as the difference between a stock return and the CRSP value-weighted market index.	CRSP, Compustat
CBOP/AT (Ball, Gerakos, Linnainmaa, and Nikolaev, 2016)	Cash-based operating profits divided by total assets. It is taken as missing for stocks with missing market value of equity, book-to-market, or total assets. SIC code must be between 6000 and 6999.	Open Asset Pricing (CZ, 2022)
Float-adjusted market capitalization	Proprietary float-adjusted market capitalization measure used by Russell to determine index weights.	FTSE Russell
Idiosyncratic volatility (Ang et al., 2006)	Standard deviation of residuals from Fama-French three factor model regressions using the past year of daily data. Require at least 200 non-missing observations.	CRSP, Ken French's Data Library

Table IA.1: Variable definitions. *(continued)*

Variable	Description	Source
IRRX (Dai et al., 2024)	Announcement-adjusted industry-relative reversals are computed as prior month's stock return less the value-weighted return to the Fama and French 49 industry, and less the CAR in that month.	CRSP, Compustat, Ken French's Data Library
Jump tail risk (Bollerslev and Todorov, 2011)	Appendix A.2 provides detailed steps on how we obtain model-free risk-neutral jump tail measures.	OptionMetrics
ln(B/M) (Fama and French, 1992)	The book-to-market ratio is the natural logarithm of the book value of a stock's common equity (ceq) divided by the firm's equity market value. We use market equity at the end of the calender year in which the firm's fiscal year ends.	Open Asset Pricing (CZ, 2022)
Liquidity beta	Sensitivity to aggregate liquidity is measured as the slope coefficient from a regression of monthly excess stock returns on the Pástor and Stambaugh liqudidity factor and the market factor. We require at least 10 non-missing observations.	CRSP, Robert Stambaugh's Home Page
Mispricing score (Stambaugh et al., 2012)	The monthly stock-level mispricing measure is a combined ranking of 11 anomaly variables.	Robert Stambaugh's Home Page
Momentum (Jegadeesh and Titman, 1993)	Momentum is the past cumulative stock return from months t-12 to t-1.	(CZ, 2022)
Passive ETF ownership (PO)	Ownership in firm i by ETF j in month t is defined as the dollar value of shares held by ETF j divided by the the total market capitalization of firm i. Total market capitalization is computed following Appel, Gormley, and Keim (2016) as the sum of market capitalization of each class of common stock associated with firm i. The dollar value of shares held by ETF j is taken from the most recent quarterly investment report. ETF ownership of each firm i is aggregated across all ETFs.	Thomson-Reuters Global Ownership, CRSP
REV (Jegadeesh, 1990)	Short-term reversal (REV) is the prior month's stock return.	CRSP
Size (Banz, 1981)	Size is the natural logarithm of market capitalization as of prior June. Market capitalization is computed as the absolute value of the number of shares outstanding times the last non-missing price in that month. It is taken as missing if either the number of shares outstanding or the share price are zero.	CRSP
Short-term reversal beta	Short-term reversal beta is the slope coefficient in a regression of daily excess returns on the short-term reversal factor. We use a window size of one year and require at least 200 non-missing observations.	CRSP, Ken French's Data Library

Table IA.1: Variable definitions. *(continued)*

Variable	Description	Source
Total market capitalization	This measure is a proxy for the proprietary Russell market value used to determine index membership. Code provided by Ben-David et al. (2019) uses standard stock databases.	Ben-David et al. (2019)
Turnover	Turnover in a given month is measured as the sum of the trading volumes during that month divided by the number of publicly held shares.	CRSP
UMC sentiment beta	We regress monthly excess stock returns on changes in UMC sentiment (orthogonalized with respect to the Fama-French three factor model) and the market factor. We require at least 10 non-missing observations.	CRSP, FRED
Volatility	Volatility in a given month is measured as the standard deviation of daily returns.	CRSP

Table IA.2: Summary statistics for the Russell sample. This table presents summary statistics of our dependent variables used in the index experiment. The sample spans 300 firms on each side of each Russell 1000/2000 index cutoff. Averages are taken across time and across firms. Data are annual from 2000 to 2021.

	N	Mean	SD	P25	Median	P75
Ownership structure						
Passive ETF ownership (%)	19,708	9.00	6.60	3.60	7.90	13.20
Passive ownership (%)	19,708	11.60	8.20	4.80	10.20	17.00
Variance decomposition						
Return variance (p.a., %)	17,197	9.20	13.80	2.80	5.20	9.90
Noise share (%)	17,197	15.1	9.1	8.8	12.7	18.8
Firm-specific information share (%)	17,197	62.7	16.7	51.9	65.0	75.5
Market information share (%)	17,197	22.3	14.8	10.4	20.4	32.1
Private information share (%)	17,197	31.6	16.4	18.9	30.4	43.1
Public information share (%)	17,197	31.1	13.7	21.0	30.4	40.3
Liquidity and sentiment						
Average bid-ask spread (%)	17,349	0.2	0.4	0.0	0.1	0.1
Liquidity beta	17,868	0.10	1.10	-0.40	0.10	0.70
Reversal beta	17,890	0.5	0.6	0.1	0.4	0.7
Idiosyncratic volatility (%)	17,890	2.10	1.20	1.30	1.70	2.50
AAII sentiment beta	17,868	0.00	1.00	-0.40	0.00	0.50
Baker and Wurgler sentiment beta	17,868	-0.00	0.40	-0.20	-0.00	0.20
UMC consumer sentiment beta	17,868	-0.00	0.90	-0.40	-0.00	0.40
Average mispricing score	12,347	49.8	12.2	41.1	49.2	57.7
Tail risk						
Jump risk (left tail, %)	18,620	25.7	18.3	14.2	20.5	31.1
Jump risk (right tail, %)	18,620	23.2	19.3	11.2	17.4	28.2
# daily return < -10%	17,368	1.4	3.6	0.0	0.0	1.0
# daily return > 10%	17,368	1.7	3.8	0.0	0.0	2.0

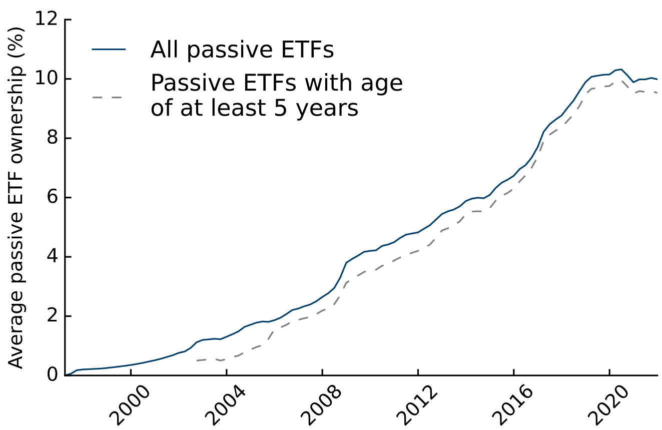


Figure IA.1: Growth of passive ETF ownership (excluding recently launched funds). This figure shows the (cross-sectional) average fraction of a stock's total market capitalization held by all passive ETFs in our sample (solid blue line). The dashed grey line depicts average PO using only holdings of ETFs that have been launched at least five years before that time. Data are displayed at the quarterly frequency and cover all U.S.-based common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

Table IA.3: Cross-sectional regressions to predict returns. This table shows average slopes (×100) and t-statistics from Fama and MacBeth cross-sectional regressions. Monthly regressions are estimated using weighted least squares (WLS) with market capitalization as weight. t-statistics are adjusted for heteroscedasticity and autocorrelation (Newey and West, 1987). Our regressions of monthly excess returns on various interactions and control variables are of the following form:

$$r^e_{i,t+1} = \alpha_t + \beta_{1,t}\text{REV}_{i,t} + \beta_{2,t}\text{Y}_{i,t} + \beta_{3,t}\text{REV}_{i,t} \times \text{Y}_{i,t} + \gamma_t\text{Controls}_{i,t} + \varepsilon_{i,t},$$

where $\text{REV}_{i,t}$ is the stock return in the prior month. $\text{Y}_{i,t}$ is either the cross-sectional quintile of the percentage of a stock's total market capitalization held by passive ETFs (PO), the total volume of trades scaled by the number of shares outstanding (TO), the standard deviation of daily stock returns within the previous month (Vola.), or the natural logarithm of a stock's market capitalization as of prior June (Size). Controls are beta, size, log book-to-market, cash-based operating profits-to-lagged assets, asset growth, and momentum. Independent variables are winsorized at the 5% and 95% levels and then standardized by their cross-sectional mean and standard deviation. Interaction terms are the product of winsorized and standardized variables. The sample excludes firms with negative book equity. Data are monthly and cover all U.S. common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
REV	-0.11	0.48	0.09	-0.34	-0.51	-0.16	-0.26	0.16	0.14	0.14
	(-0.9)	(2.22)	(0.53)	(-2.97)	(-4.73)	(-1.46)	(-2.71)	(0.49)	(0.55)	(0.51)
PO		-0.04	-0.03							-0.05
		(-0.8)	(-0.84)							(-1.19)
REV x PO		-0.17	-0.11							-0.1
		(-3.98)	(-3.21)							(-3.28)
TO				0.1	0.07					0.17
				(0.7)	(0.76)					(2.01)
REV x TO				0.3	0.29					0.27
				(4.66)	(4.82)					(4.86)
Volatility						-0.01	-0.25			-0.34
						(-0.05)	(-1.78)			(-2.68)
REV x Vola.						0.24	0.29			0.03
						(3.06)	(3.94)			(0.42)
Size								-0.02	-0.05	-0.07
								(-0.3)	(-1.12)	(-1.32)
REV x Size								-0.04	-0.05	-0.03
								(-1.02)	(-1.65)	(-1.06)
Controls	No	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Adj. R^2 (%)	2.67	3.85	15.17	6.45	15.95	6.88	15.88	5.93	15.12	17.59
N	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7

Table IA.4: Cross-sectional regressions to predict returns. This table shows average slopes ($\times 100$) and t-statistics from Fama and MacBeth cross-sectional regressions. Monthly regressions are estimated using weighted least squares (WLS) with market capitalization as weight. t-statistics are adjusted for heteroscedasticity and autocorrelation (Newey and West, 1987). Our regressions of monthly excess returns on various interactions and control variables are of the following form:

$$r^e_{i,t+1} = \alpha_t + \beta_{1,t}\mathrm{PO}_{i,t} + \beta_{2,t}\mathrm{Y}_{i,t} + \beta_{3,t}\mathrm{Y}_{i,t} \times \mathrm{PO}_{i,t} + \gamma_t \mathrm{Controls}_{i,t} + \varepsilon_{i,t},$$

where $\mathrm{PO}_{i,t}$ is the cross-sectional quintile of the percentage of a stock's total market capitalization held by passive ETFs. $\mathrm{Y}_{i,t}$ is either the previous month's return (REV), the announcement-adjusted return (REVX), the industry-relative return (IRR), or the announcement-adjusted industry-relative return (IRRX). Controls are beta, size, log book-to-market, cash-based operating profits-to-lagged assets, asset growth, and momentum. Independent variables are winsorized at the 5% and 95% levels and then standardized by their cross-sectional mean and standard deviation. Interaction terms are the product of winsorized and standardized variables. The sample excludes firms with negative book equity. Data are monthly and cover all U.S. common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
PO	-0.04	-0.05	-0.05	-0.04	-0.04	-0.05	-0.04	-0.05
	(-0.8)	(-1.27)	(-0.87)	(-1.24)	(-0.58)	(-1.37)	(-0.63)	(-1.37)
REV	0.47	0.08						
	(2.21)	(0.5)						
REV x PO	-0.17	-0.11						
	(-3.98)	(-3.12)						
REVX			0.43	0.04				
			(1.97)	(0.24)				
REVX x PO			-0.18	-0.12				
			(-4.02)	(-3.18)				
IRR					0.41	0.1		
					(2.01)	(0.63)		
IRR x PO					-0.16	-0.1		
					(-3.44)	(-2.83)		
IRRX							0.33	0.06
							(1.62)	(0.35)
IRRX x PO							-0.16	-0.11
							(-3.42)	(-2.89)
Controls	No	Yes	No	Yes	No	Yes	No	Yes
Adj. R^2 (%)	3.85	14.69	3.95	14.73	2.42	14.22	2.42	14.24
N	3025.98	3025.98	3025.98	3025.98	3025.98	3025.98	3025.98	3025.98

Table IA.5: Impact of passive ownership on sentiment betas excluding recently launched ETFs. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on sentiment betas and average mispricing scores. Specifically, we estimate

$$\begin{aligned} Y_{i,t} \quad = \quad & \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ & + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. The dependent variable is a stock's beta with respect to changes in an aggregate sentiment measure or its mispricing score according to Stambaugh et al. (2012). Measures of aggregate sentiment are the American Association of Individual Investors (AAII) sentiment survey, the Baker and Wurgler (2007) sentiment index, and UMC consumer sentiment, respectively. AAII sentiment is a monthly average of the bull-bear-spread, defined as the percentage of bullish investors minus the percentage of bearish investors. Sentiment betas are from a regression of monthly excess returns on orthogonalized changes in sentiment and the market factor. Orthogonalization is performed with respect to the Fama-French three factor model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	AAII sentiment beta	BW sentiment beta	UMC sentiment beta	Mispricing score
$\widehat{PO}_{5y}$	0.49**	0.45**	0.58**	0.6**
	(2.04)	(2.03)	(2.04)	(2.09)
Bandwidth	300	300	300	300
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	17868	17868	17868	12347
R-squared (%)	0.29	0.26	0.19	0.90

Table IA.6: Impact of index assignment on passive ownership. This table presents the first stage of an instrumental variable regression of passive ownership on membership in the Russell 2000 index and further controls. Specifically, we estimate

$$\begin{aligned} Passive_{i,t} &= \eta + \lambda R2000_{i,t} + \sum_{n=1}^{N} \chi_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $Passive_{i,t}$ is the percentage market capitalization of firm i owned by passive funds at the end of September in year t. $R2000_{i,t}$ is our instrument and indicates a stock's assignment to the Russell 2000 at the end of June in year t. $Mktcap_{i,t}$ controls for end of May market capitalization computed using the method of Ben-David et al. (2019) and $Float_{i,t}$ is float-adjusted market capitalization in June used by Russell to determine index weights. To account for Russell's banding rule Appel et al. (2019) add three further controls: an indicator for a stock's membership in the Russell 2000 in the previous year ($R2000_{i,t-1}$), an indicator for whether a stock falls within a certain range close to the Russell cutoff ($Band_{i,t}$), and their interaction. The dependent variable is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Passive ownership			
	(1)	(2)	(3)	(4)
R2000	0.1***	0.11***	0.12***	0.14***
	(5.0)	(5.9)	(6.75)	(7.97)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	14787	19707	24107	28506
R-squared (%)	0.45	0.42	0.38	0.45

Table IA.7: Impact of passive ownership on variance components. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ownership on return variance component shares. Specifically, we estimate

$$\begin{aligned} VarianceShare_{i,t} \quad &= \quad \alpha + \beta \widehat{Passive}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &\quad + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{Passive}_{i,t}$ is the percentage market capitalization of firm i owned by passive funds at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $VarianceShare_{i,t}$ are the components obtained from the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1;^{**}p < 0.05;^{***}p < 0.01$.

		Variance shares			FirmInfo	
Dependent variable =	Variance	Noise	FirmInfo	MktInfo	PrivateInfo	PublicInfo
$\widehat{Passive}$	15.88***	7.96***	-11.3***	3.34	-17.97***	6.67*
	(4.09)	(3.65)	(-2.84)	(0.94)	(-4.31)	(1.91)
Bandwidth	300	300	300	300	300	300
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17197	17197	17197	17197	17197	17197
R-squared (%)	3.02	0.34	1.76	2.60	0.20	2.12

Table IA.8: Cross-sectional regressions to predict returns. This table shows average slopes (×100) and t-statistics from Fama and MacBeth cross-sectional regressions. Monthly regressions are estimated using weighted least squares (WLS) with market capitalization as weight. t-statistics are adjusted for heteroscedasticity and autocorrelation (Newey and West, 1987). Our regressions of monthly excess returns on various interactions and control variables are of the following form:

$$r^e_{i,t+1} = \alpha_t + \beta_{1,t}\mathrm{REV}_{i,t} + \beta_{2,t}\mathrm{Y}_{i,t} + \beta_{3,t}\mathrm{REV}_{i,t} \times \mathrm{Y}_{i,t} + \gamma_t\mathrm{Controls}_{i,t} + \varepsilon_{i,t},$$

where $\mathrm{REV}_{i,t}$ is the stock return in the prior month. $\mathrm{Y}_{i,t}$ is either the cross-sectional quintile of the percentage of a stock's total market capitalization held by passive funds (Passive), the total volume of trades scaled by the number of shares outstanding (TO), the standard deviation of daily stock returns within the previous month (Vol.), or the natural logarithm of a stock's market capitalization as of prior June (Size). Controls are beta, size, log book-to-market, cash-based operating profits-to-lagged assets, asset growth, and momentum. Independent variables are winsorized at the 5% and 95% levels and then standardized by their cross-sectional mean and standard deviation. Interaction terms are the product of winsorized and standardized variables. The sample excludes firms with negative book equity. Data are monthly and cover all U.S. common stocks on NYSE, AMEX, and NASDAQ from June 1997 to December 2021.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
REV	-0.11	0.52	0.16	-0.34	-0.51	-0.16	-0.26	0.16	0.14	0.12
	(-0.9)	(2.36)	(0.9)	(-2.97)	(-4.73)	(-1.46)	(-2.71)	(0.49)	(0.55)	(0.44)
Passive		-0.03	-0.01							-0.02
		(-0.48)	(-0.13)							(-0.57)
REV x Passive		-0.18	-0.13							-0.11
		(-4.1)	(-3.59)							(-3.34)
TO				0.1	0.07					0.17
				(0.7)	(0.76)					(1.99)
REV x TO				0.3	0.29					0.27
				(4.66)	(4.82)					(4.84)
Volatility						-0.01	-0.25			-0.34
						(-0.05)	(-1.78)			(-2.67)
REV x Vol.						0.24	0.29			0.02
						(3.06)	(3.94)			(0.37)
Size								-0.02	-0.05	-0.06
								(-0.3)	(-1.12)	(-1.3)
REV x Size								-0.04	-0.05	-0.03
								(-1.02)	(-1.65)	(-0.9)
Controls	No	No	Yes	No	Yes	No	Yes	No	Yes	Yes
Adj. R2 (%)	2.67	3.81	15.17	6.45	15.95	6.88	15.88	5.93	15.12	17.6
N	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7	3040.7

Table IA.9: Impact of passive ETF ownership on bid-ask spreads. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on bid-ask spreads. Specifically, we estimate

$$\begin{aligned}\overline{baspread}_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},\end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $\overline{baspread}_{i,t}$ is the average percentage bid-ask spread in the 12 months after index reconstitution. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Average bid-ask spread			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	1.03***	0.9***	0.77***	0.71***
	(4.05)	(4.56)	(4.48)	(4.98)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12973	17349	21275	25295
R-squared (%)	3.50	2.07	2.32	2.54

Table IA.10: Impact of passive ETF ownership on exposure to aggregate liquidity. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on liquidity betas. Specifically, we estimate

$$\begin{aligned}\beta^{PSLIQ}_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},\end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $\beta^{PSLIQ}_{i,t}$ is the slope coefficient from a regression of monthly excess stock returns on the Pástor and Stambaugh (2003) liquidity factor. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Pástor and Stambaugh liquidity factor beta			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	0.95***	0.92***	0.95***	0.8***
	(3.35)	(3.67)	(4.15)	(4.04)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	13408	17868	21911	25960
R-squared (%)	0.31	0.37	0.37	0.35

Table IA.11: Impact of passive ETF ownership on short-term reversal. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on short-term reversal betas. Specifically, we estimate

$$\begin{aligned}\beta_{i,t}^{STREV} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},\end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $\beta_{i,t}^{STREV}$ is the slope coefficient from a regression of the short-term reversal factor on daily excess stock returns in the year after index reconstitution. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Short-term reversal beta			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	0.6***	0.69***	0.69***	0.65***
	(2.85)	(3.69)	(4.0)	(4.29)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	13426	17890	21939	25995
R-squared (%)	0.36	0.42	0.50	0.48

Table IA.12: Impact of passive ETF ownership on idiosyncratic volatility. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on idiosyncratic volatility. Specifically, we estimate

$$\begin{aligned} IdioVol_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $IdioVol_{i,t}$ is the standard deviation of residuals from a Fama-French three factor model regression in the year after index reconstitution. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Idiosyncratic volatility			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	0.44*	0.62***	0.79***	0.91***
	(1.77)	(2.79)	(3.88)	(5.09)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	13426	17890	21939	25995
R-squared (%)	5.51	6.03	6.85	7.59

Table IA.13: Impact of passive ETF ownership on risk-neutral jump tail measures. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on tail risk. Specifically, we estimate

$$\begin{aligned} Y_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $Y_{i,t}$ is the option-implied tail jump risk measure following Bollerslev and Todorov (2011). Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	$LT^{\mathbb{Q}} \approx \frac{e^{r_{t,T}(T-t)}P_t(K)}{(T-t)F_t}$			$RT^{\mathbb{Q}} \approx \frac{e^{r_{t,T}(T-t)}C_t(K)}{(T-t)F_t}$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{PO}$	19.18***	22.01***	21.83***	19.4***	22.63***	22.58***
	(4.7)	(5.92)	(6.78)	(4.29)	(5.49)	(6.34)
Bandwidth	300	400	500	300	400	500
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	18620	22652	26706	18620	22652	26706
R-squared (%)	9.82	11.44	13.50	10.33	12.01	14.11

Table IA.14: Impact of passive ETF ownership on the number of extreme returns. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on extreme price movements. Specifically, we estimate

$$\begin{aligned} Y_{i,t} =& \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $Y_{i,t}$ is the number of daily returns in the year after index reconstitution smaller (larger) than the threshold of -10% (10%). Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1; ^{**}p < 0.05; ^{***}p < 0.01$.

Dependent variable =	$count(r_{d,i,t} < -10\%)$			$count(r_{d,i,t} > 10\%)$		
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{PO}$	3.35***	3.08***	3.16***	2.28**	2.6***	2.89***
	(3.9)	(4.1)	(4.87)	(2.59)	(3.38)	(4.37)
Bandwidth	300	400	500	300	400	500
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17368	21299	25319	17368	21299	25319
R-squared (%)	1.58	1.89	2.00	2.65	3.02	3.17

Table IA.15: Impact of passive ETF ownership on return variance. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on the share of firm-specific information in return variance. Specifically, we estimate

$$\begin{aligned} Variance_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $FirmInfoShare_{i,t}$ is the sum of public and private firm-specific information components in the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Return variance (%)			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	8.79**	12.79***	12.86***	13.52***
	(2.58)	(4.09)	(4.75)	(5.76)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12865	17197	21091	25083
R-squared (%)	2.75	3.02	3.49	3.70

Table IA.16: Impact of passive ETF ownership on the share of noise in return variance. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on the share of noise in return variance. Specifically, we estimate

$$NoiseShare_{i,t} = \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $NoiseShare_{i,t}$ is the noise component in the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Noise share (%)			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	8.09***	6.41***	6.61***	6.02***
	(4.03)	(3.65)	(4.23)	(4.45)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12865	17197	21091	25083
R-squared (%)	0.40	0.34	0.32	0.31

Table IA.17: Impact of passive ETF ownership on the share of firm-specific information in return variance. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on the share of firm-specific information in return variance. Specifically, we estimate

$$\begin{aligned} FirmInfoShare_{i,t} = & \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ & + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $FirmInfoShare_{i,t}$ is the sum of public and private firm-specific information components in the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Firm-specific information share (%)			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	-11.92***	-9.1***	-6.87**	-4.42*
	(-3.28)	(-2.84)	(-2.41)	(-1.77)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12865	17197	21091	25083
R-squared (%)	1.66	1.76	1.96	2.09

Table IA.18: Impact of passive ETF ownership on the share of market information in return variance. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on the share of market information in return variance. Specifically, we estimate

$$\begin{aligned} MktInfoShare_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $MktInfoShare_{i,t}$ is the market information component in the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1;^{**}p < 0.05;^{***}p < 0.01$.

Dependent variable =	Market information share (%)			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	3.83	2.69	0.26	-1.6
	(1.18)	(0.94)	(0.1)	(-0.72)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12865	17197	21091	25083
R-squared (%)	2.37	2.60	2.70	2.75

Table IA.19: Impact of passive ETF ownership on the share of private firm-specific information in return variance. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on the share of private firm-specific information in return variance. Specifically, we estimate

$$PrivateInfoShare_{i,t} = \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} + \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $PrivateInfoShare_{i,t}$ is the private firm-specific information component in the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Private firm-specific information share (%)			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	-15.8***	-14.47***	-12.63***	-10.4***
	(-4.08)	(-4.31)	(-4.26)	(-4.01)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12865	17197	21091	25083
R-squared (%)	0.21	0.20	0.19	0.18

Table IA.20: Impact of passive ETF ownership on the share of public firm-specific information in return variance. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on the share of public firm-specific information in return variance. Specifically, we estimate

$$\begin{aligned} PublicInfoShare_{i,t} = &\alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $PublicInfoShare_{i,t}$ is the public firm-specific information component in the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Public firm-specific information share (%)			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	3.88	5.37*	5.76**	5.98***
	(1.24)	(1.91)	(2.33)	(2.82)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	12865	17197	21091	25083
R-squared (%)	2.08	2.12	2.35	2.75

Table IA.21: Impact of passive ETF ownership on stock return sensitivity to changes in sentiment. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on sentiment betas. Specifically, we estimate

$$\begin{aligned} Y_{i,t} =& \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $Y_{i,t}$ is a stock's beta w.r.t. changes in the American Association of Individual Investors (AAII) sentiment survey or the Baker and Wurgler (2007) sentiment index. AAII sentiment is a monthly average of the bull-bear-spread, defined as the percentage of bullish investors minus the percentage of bearish investors. We first orthogonalize changes in sentiment with respect to the Fama-French three factor model using a full-sample time-series regression. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	AAII sentiment beta			BW sentiment index beta		
	(1)	(2)	(3)	(4)	(5)	(6)
$\widehat{PO}$	0.46**	0.5**	0.44**	0.42**	0.33*	0.23
	(2.04)	(2.46)	(2.48)	(2.03)	(1.75)	(1.38)
Bandwidth	300	400	500	300	400	500
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17868	21911	25960	17868	21911	25960
R-squared (%)	0.29	0.29	0.26	0.26	0.25	0.24

Table IA.22: Impact of passive ETF ownership on UMC sentiment beta. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on a stock's exposure to changes in UMC consumer sentiment. Specifically, we estimate

$$\beta^{\text{UMC}}_{i,t} = \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N}\lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} + \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $\beta^{\text{UMC}}_{i,t}$ is a stock's beta w.r.t. changes in the UMC consumer sentiment index. We first orthogonalize changes in sentiment with respect to the Fama-French three factor model using a full-sample time-series regression. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2016. $^*p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	UMC sentiment beta			
	(1)	(2)	(3)	(4)
$\widehat{ETF}$	0.68**	0.54**	0.34	0.22
	(2.22)	(2.04)	(1.45)	(1.08)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	13408	17868	21911	25960
R-squared (%)	0.24	0.19	0.12	0.11

Table IA.23: Impact of passive ETF ownership on mispricing. This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on stock mispricing. Specifically, we estimate

$$\begin{aligned}\overline{misp}_{i,t} =& \alpha + \beta\widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n(\ln(Mktcap_{i,t}))^n + \sigma\ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3(Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},\end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $\overline{misp}_{i,t}$ is the Stambaugh et al. mispricing measure, averaged across 12 months following index reconstitution. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership and the dependent variable are scaled by their sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2016. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Average mispricing score			
	(1)	(2)	(3)	(4)
$\widehat{PO}$	0.66**	0.56**	0.49**	0.39*
	(2.07)	(2.09)	(2.04)	(1.67)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	9281	12347	15318	18338
R-squared (%)	0.74	0.90	1.17	1.68

Table IA.24: Impact of index assignment on passive ETF ownership (using float-adjusted market capitalization ranks). This table presents the first stage of an instrumental variable regression of passive ETF ownership on membership in the Russell 2000 index and further controls. Specifically, we estimate

$$PO_{i,t} = \eta + \lambda R2000_{i,t} + \sum_{n=1}^{N} \chi_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} + \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t},$$

where $PO_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t. $R2000_{i,t}$ is our instrument and indicates a stock's assignment to the Russell 2000 at the end of June in year t. $Mktcap_{i,t}$ controls for end of May market capitalization computed using the method of Ben-David et al. (2019) and $Float_{i,t}$ is float-adjusted market capitalization in June used by Russell to determine index weights. To account for Russell's banding rule Appel et al. (2019) add three further controls: an indicator for a stock's membership in the Russell 2000 in the previous year ($R2000_{i,t-1}$), an indicator for whether a stock falls within a certain range close to the Russell cutoff ($Band_{i,t}$), and their interaction. The dependent variable is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Stocks are ranked based on end of June float-adjusted market capitalization. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$.

Dependent variable =	Passive ETF ownership			
	(1)	(2)	(3)	(4)
R2000	0.18***	0.16***	0.17***	0.18***
	(8.26)	(8.41)	(9.59)	(10.19)
Bandwidth	200	300	400	500
Polynomial order, N	3	3	3	3
Float control	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes
Observations	14789	19709	24108	28508
R-squared (%)	0.39	0.30	0.29	0.30

Table IA.25: Impact of passive ETF ownership on variance components (using float-adjusted market capitalization ranks). This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on return variance component shares. Specifically, we estimate

$$\begin{aligned} VarianceShare_{i,t} =& \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $VarianceShare_{i,t}$ are the components obtained from the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $^{*}p < 0.1; ^{**}p < 0.05; ^{***}p < 0.01$.

		Variance shares			FirmInfo	
Dependent variable =	Variance	Noise	FirmInfo	MktInfo	PrivateInfo	PublicInfo
$\widehat{PO}$	7.28***	6.09***	-8.01***	1.93	-10.2***	2.18
	(3.29)	(3.88)	(-2.82)	(0.76)	(-3.42)	(0.9)
Bandwidth	300	300	300	300	300	300
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17193	17193	17193	17193	17193	17193
R-squared (%)	1.78	0.31	1.14	1.62	0.12	0.98

Table IA.26: Impact of passive ETF ownership on variance components (using $t = 60$ instead of $t = 15$). This table presents the second stage of an instrumental variable estimation used to identify the causal effect of passive ETF ownership on return variance component shares. Specifically, we estimate

$$\begin{aligned} VarianceShare_{i,t} =& \alpha + \beta \widehat{PO}_{i,t} + \sum_{n=1}^{N} \lambda_n (\ln(Mktcap_{i,t}))^n + \sigma \ln(Float_{i,t}) + \phi_1 Band_{i,t} \\ &+ \phi_2 R2000_{i,t-1} + \phi_3 (Band_{i,t} \times R2000_{i,t-1}) + \delta_t + \varepsilon_{i,t}, \end{aligned}$$

where $\widehat{PO}_{i,t}$ is the percentage market capitalization of firm i owned by passive ETFs at the end of September in year t, instrumented using exogenous assignment to the Russell 2000 index. $VarianceShare_{i,t}$ are the components obtained from the Brogaard et al. (2022) return variance decomposition model. Controls are end-of-May market capitalization, float-adjusted market capitalization at the end of June, and additional variables to account for Russell's banding rule following Appel et al. (2019). Ownership is scaled by its sample standard deviation. t-statistics are in parentheses and based on clustered standard errors at the firm level. Bandwidths indicate the number of firms around each Russell cutoff included in our sample. Data are annual and cover reconstitutions from 2000 to 2020. $*p < 0.1; **p < 0.05; ***p < 0.01$.

		Variance shares			FirmInfo	
Dependent variable =	Variance	Noise	FirmInfo	MktInfo	PrivateInfo	PublicInfo
$\widehat{PO}$	13.0***	6.09***	-8.1**	2.01	-13.73***	5.63**
	(4.15)	(3.5)	(-2.52)	(0.7)	(-4.1)	(1.99)
Bandwidth	300	300	300	300	300	300
Polynomial order, N	3	3	3	3	3	3
Float control	Yes	Yes	Yes	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	17198	17198	17198	17198	17198	17198
R-squared (%)	3.02	0.30	1.90	2.70	0.21	2.22