The Review of Financial Studies

Exchange-Traded Funds and Real Investment

Constantinos Antoniou
University of Warwick, UK

Frank Weikai Li
Lee Kong Chian School of Business, Singapore Management University, Singapore

Xuewen Liu
The University of Hong Kong, Hong Kong

Avanidhar Subrahmanyam
University of California, Los Angeles, USA

Chengzhu Sun
School of Accounting and Finance, the Hong Kong Polytechnic University, Hong Kong

We investigate the link between exchange-traded funds and real investment. Cross-sectionally, higher ETF ownership is associated with an increased sensitivity of real investment to Tobin's q and a heightened ability of stock returns to forecast future earnings. Inclusion of stocks in industry ETFs enhances investment-q sensitivity and implies greater incorporation of earnings information into prices prior to public releases. Greater nonmarket ETF ownership leads to increased (reduced) reliance of real investment on own (peers') stock prices. Overall, the evidence is consistent with ETFs positively affecting real investment efficiency via greater flows of information. (*JEL* G14, G23, G31)

Received February 3, 2020; editorial decision May 16, 2022 by Editor Itay Goldstein. Authors have furnished an Internet Appendix, which is available on the Oxford University Press Web site next to the link to the final published paper online.

We are grateful to two anonymous referees and Itay Goldstein (the editor) for insightful and constructive feedback. We also thank Gennaro Bernile, Utpal Bhattacharya, Xin Chang, Si Cheng, Olivier Dessaint, Fangjian Fu, Claire Hong, Laurent Fresard, Jasmin Gider, Roman Kozhan, Nitin Kumar, Hao Liang, Roger Loh, Abhiroop Mukherjee, Kasper Nielsen, Clemens Otto, Sophie Shive, Baolian Wang, Xin Wang, Chishen Wei, Junyu (Tiffany) Wang, Dong Yan, and Xiaoyun Yu and seminar participants at the 2018 China International Conference in Finance, the 2019 ISB Summer Research Conference, the SHoF Conference on Financial Markets and Corporate Decisions, the 2018 Asia Quant Conference, HKUST, Singapore Management University, and Nanyang Technological University for helpful comments. This paper is based on an original paper titled "The Real Effects of Exchange-Traded Funds" by Weikai Li, Xuewen Liu, and Chengzhu Sun and absorbs the related paper entitled "ETF Ownership and Corporate Investment" written independently by Constantinos Antoniou, Avanidhar Subrahmanyam, and Onur Tosun. Supplementary data can be found on *The Review of Financial Studies* web site. Send correspondence to Avanidhar Subrahmanyam, asubrahm@anderson.ucla.edu.

The Review of Financial Studies 36 (2023) 1043–1093

https://doi.org/10.1093/rfs/hhac043 Advance Access publication July 20, 2022

The exchange-traded fund (ETF) industry has grown spectacularly over recent years.[1] While popular among investors, ETFs can increase systemic risk and induce nonfundamental volatility, as well as excess comovement, recent evidence suggests (Ben-David, Franzoni, and Moussawi 2018; Da and Shive 2018). Debates about the benefits and the potential destabilizing effects of ETFs are in their early stages, and a fuller understanding of the overall implications of ETFs is critical for regulators.

So far researchers have largely focused on studying the effects of ETFs on the market (informational) efficiency of the underlying securities. The evidence is mixed. On the one hand, Israeli, Lee, and Sridharan (2017), among others, find that firms that are widely held by ETFs appear to experience a decrease in informational efficiency with regard to firm-specific information. On the other hand, Glosten, Nallareddy, and Zou (2021) and Bhojraj, Mohanram, and Zhang (2020) find that ETF activity facilitates the timely incorporation of earnings information into stock prices.

While the effects of ETFs on market efficiency are clearly an important avenue to research, a more complete examination of ETFs' impact should include the study of the links between ETFs and real investment. Indeed, as Bond, Edmans, and Goldstein (2012) propose, the evaluation of price efficiency should be in terms of prices' usefulness for real decisions, beyond the degree to which they forecast cash flows. The potential impact of rising ETF ownership on the allocational role of asset prices is also a concern among market participants.[2] In this paper, we contribute to the debate surrounding ETFs by studying the relation between ETFs and corporate investment policies. We are not aware of prior research that explores this link.

There are reasons to believe that the link between ETFs and the efficiency of real investment can go either way. For example, Ben-David, Franzoni, and Moussawi (2018) show that nonfundamental volatility increases with ETF ownership. This result suggests that ETF ownership should reduce the sensitivity of real investments to securities' market values. On the other hand, ETF ownership might increase price informativeness about cash flow shocks. This happens if the number of factor-informed traders increases upon introduction of the ETF basket security, and these traders trade both ETFs and the underlying security (Subrahmanyam 1991). The heightened informativeness should be an increasing function of ETF ownership.[3] Thus,

[1] According to the Investment Company Institute, as of April 2017, $2.84 trillion was being managed by 1,751 ETFs in the U.S. market. As of 2015, over 10% of the market capitalization and about 35% of the trading volume of securities traded on U.S. stock exchanges are attributable to ETFs (Ben-David, Franzoni, and Moussawi 2017).

[2] In a research report entitled "The Silent Road to Serfdom: Why Passive Investing is Worse than Marxism," strategists from the research and brokerage firm Sanford Bernstein argue that a capitalist system with indexed investing via ETFs may be less desirable than a centrally planned economy, where governments direct all real investment. See, for example, commentary by Malkiel (2016).

[3] We justify this observation as follows: First, ETFs attract more noise trading, subsidizing more factor-informed trading. Furthermore, high ETF ownership in a stock allows for the timely incorporation of systematic information

when ownership by ETFs increases, the firm's investment might be more responsive to its own stock price. We build a simple model with information asymmetry to formalize the latter intuition and also develop additional testable implications.[4]

We test the contrasting hypotheses suggested by earlier literature and the central result of our model using a large sample of U.S. equity ETFs. We find that higher ETF ownership is associated with a greater sensitivity of real investment to Tobin's q. The economic magnitude is nontrivial: one interquartile increase in ETF ownership increases investment-q sensitivity by 8.3%. To address endogeneity concerns, we use BlackRock's acquisition of iShares from Barclays at the end of 2009 as an exogenous shift in ETF ownership (Zou 2019). Because of this event, iShares ETFs experienced a significant increase in fund flows relative to non-iShares ETFs. Our instrument is a dummy that takes a value of one for stocks with above-median iShares ETF ownership measured before the acquisition. Using a set of treatment and matched control firms, we provide evidence consistent with the notion that ETF ownership causally affects the investment-q sensitivity of firms.

The findings in recent literature suggest that informed trading is linked to ETFs that track specific industry sectors (e.g., Bhojraj, Mohanram, and Zhang 2020; Huang, O'Hara, and Zhong 2021). Moreover, since market index prices are readily available, we would expect a learning channel to primarily operate via nonmarket ETFs that do not mimic such indexes. Motivated by these observations, we split ETF ownership into that by market and nonmarket ETFs, and reestimate our baseline model. We find that higher ownership by nonmarket ETFs increases the sensitivity of investments to stock prices, whereas ownership by market ETFs bears no relationship to this sensitivity. We supplement this analysis with an identification strategy based on stocks' inclusion in industry ETFs (Huang, O'Hara, and Zhong 2021). We find that the investment sensitivity to q increases for firms that are added for the first time to an industry ETF, relative to a control sample. Further, we find that the inclusion event increases the sensitivity of investments to the common component of q. We also use the return around earnings announcements as an inverse proxy for information conveyed by market prices prior to the announcement, and find that inclusion in industry ETFs attenuates the common component of this return. In contrast, the inclusion of a firm in a market index ETF does not alter the sensitivity of its real investment to q, nor does it influence the response of stock prices to earnings surprises. These findings collectively suggest that ownership by nonmarket

via arbitrage forces, making the stock more desirable to factor-informed traders who wish to trade both the ETF and individual stocks. As we will show, the heightened number of factor-informed traders can also increase the incentives for firm-specific informed traders to collect information.

[4] In our model, the ETF directly increases the number of factor-informed traders. This has an indirect effect of increasing firm-specific informed traders because of heightened competition between the factor-informed traders (as in Subrahmanyam 1991). The two effects reinforce each other. In our empirical work, we find the direct effect to be stronger, but we cannot rule out the indirect effect.

ETFs brings fundamental information into prices, which the manager uses in real decisions.

To substantiate the learning mechanism, we also test our model's prediction regarding investment sensitivity to peers' stock prices (Foucault and Frésard 2014; Dessaint et al. 2019). Our analysis indicates that higher nonmarket ETF ownership should imply a higher (lower) sensitivity of investment to own (peers') q. The evidence is consistent with these hypotheses, which accords with the information channel. To further investigate if the information transmitted by nonmarket ETFs into stock prices is used by managers, we conduct an empirical test based on Dessaint et al. (2019) that uses mutual fund redemptions as an exogenous shock to prices. Because it is unlikely that the manager knows the component of price movements due to redemptions, evidence of real investments' dependence on this component supports managerial learning from prices. We decompose both own and peers' q into a component related to the redemptions variable, and an orthogonal component. We find that nonmarket ETF ownership enhances (reduces) investment-q sensitivity for the flow-related component of own (peers') q. This provides support for the notion that managers condition on prices to make real investment decisions.

Our model offers two additional cross-sectional predictions that we test in the paper. We predict the positive effect of nonmarket ETF ownership on investment-q sensitivity to be stronger when the precision of common information is lower. Using stocks' cash flow beta and the volatility of industry-level profitability as proxies for precision of the common factor, we find evidence consistent with this prediction. Next, our model predicts that the positive effect of nonmarket ETF ownership on investment-q sensitivity is stronger when the firm manager has more precise firm-specific information. Using the profitability of insider trades as a proxy for the precision of managerial (firm-specific) information, we find supporting evidence for this implication as well.

A natural question related to our hypothesis and empirical findings is why managers prefer to rely on own stock prices over ETF prices to learn about common information. First, there may be complementarity between cash flow components in terms of information acquisition. Specifically, more information about supply and demand components for oil, for example, could facilitate information production about transportation. Thus, while nonmarket ETFs facilitate the incorporation of common information in stock prices, these latter prices can contain additional information beyond the ETF prices. Second, as long as the ETF prices are noisy,[5] they are not perfect substitutes for stock prices. Third, we propose that stock prices are more salient for managers (Hirshleifer and Teoh 2003; Hong, Torous, and Valkanov 2007). Indeed, extracting information from dozens of ETF prices is costly under limited

5 For example, these prices may contain information about additional factors that do not affect the value of the stock or include the effects of ETF-specific noise trading.

attention; on average, a stock is held by more than 20 ETFs, and the maximum number of ETFs holding shares in a stock is more than 100 within our sample. We investigate the salience argument by proposing that we expect the positive effect of nonmarket ETFs to be stronger when the average correlation between returns on the stock and nonmarket ETFs holding the stock is low. This is because when this correlation is high, the firm is to a large extent exposed to the same common factors as the ETFs holding the stock of the firm, and therefore learning from the ETF prices is easier, and vice versa. We indeed find that the effect of nonmarket ETF ownership on investment-price sensitivity is higher when the average correlation between stock returns and those of nonmarket ETFs owning the stock is lower.

We also explore alternative explanations for our main findings. First, recent literature has documented that passive institutional ownership may improve corporate governance quality (Appel, Gormley, and Keim 2016), which could lead to heightened investment-q sensitivity. Second, firms that are held by more ETFs could have easier access to external finance and face fewer financial constraints. This could strengthen the investment-q sensitivity by allowing firms to better exploit investment opportunities. In our robustness checks, however, we find that ETFs improve investment-q sensitivity only among firms with strong corporate governance to begin with, and that measures of financial constraints are not significantly affected by firms' ETF ownership.

Our paper contributes to two strands of the literature. The first is the growing body of work on the impact of ETFs on financial markets. Several papers have argued that demand shocks transmitted from the ETFs to their underlying securities affect the pricing of the latter. Ben-David, Franzoni, and Moussawi (2018) indicate that ETF-relate arbitrage activities increase underlying stocks' volatility, and Israeli, Lee, and Sridharan (2017) propose that ETFs can lead to lower liquidity of constituent stocks. We note that our findings are not inconsistent with these papers. Ben-David, Franzoni, and Moussawi (2018) study daily stock returns. While index ETFs can increase short-term (daily) volatility of the underlying stocks, nonmarket ETFs can simultaneously improve long-run (quarterly, yearly) stock price informativeness about industry information. Israeli, Lee, and Sridharan (2017) argue that ETF ownership affects firm-specific informativeness, but they do not specifically focus on common information. Moreover, they study the lagged effects of ETFs.[6]

In other work, Bhattacharya and O'Hara (2018) show within a theoretical setting that feedback between ETFs and their constituents can cause

[6] Specifically, consider a generic time index t, and let $Ret(t)$, $ETF(t)$, $Earn(t)$, $Q(t)$, and $Inv(t)$ denote returns, ETF ownership, earnings, Tobin's q, and real investment, respectively, at t. Israeli, Lee, and Sridharan (2017) investigate the relationship between $ETF(t-1)$, $Ret(t)$, and $Earn(t+1)$ using annual data. We study the relationship between $ETF(t-1)$, $Q(t-1)$, and $Inv(t)$, similar to the specification in Foucault and Frésard (2012), who study the relationship between cross-listing status at $t-1$, $Q(t-1)$ and $Inv(t)$. We argue that as a stock market is competitive, with few entry barriers, the effect of ETF ownership on the informativeness of stock prices is contemporaneous, and thus in our models both quantities enter at the same time point $(t-1)$.

propagation of shocks unrelated to fundamentals. Subrahmanyam (1991) and Cong and Xu (2019) propose that initiation of basket securities can reduce (increase) speculators' incentives to acquire and trade on asset-specific (common) information.[7] Glosten, Nallareddy, and Zou (2021) provide evidence that ETF trading increases informational efficiency for stocks with weak information environments. Bhojraj, Mohanram, and Zhang (2020) find that sector ETFs are effective at transmitting industry information across firms. Complementing these contributions, our paper studies the link between ETFs and real investment.

Our second contribution relates to the long-standing and important debate on whether financial markets affect the real economy or are merely a sideshow. Several theory papers have proposed the managerial learning hypothesis, which posits that when speculators trade on their private signals, the stock price is useful to real decision makers.[8] The majority of empirical studies on the managerial learning hypothesis take the sensitivity of corporate investment to stock price as evidence of real feedback from financial markets viz., Chen, Goldstein, and Jiang (2007), Bakke and Whited (2010), and Foucault and Frésard (2012, 2014). In addition to the preceding studies, supporting evidence points to managerial learning by using specific settings. Luo (2005) finds that managers are more likely to cancel acquisition plans when the market's response to a deal announcement is negative. Zuo (2016) documents that recent stock price changes positively affect a manager's belief about fundamentals. Other settings use a firm's cross-listing status, the staggered enforcement of insider trading laws across countries, and changes in mandatory disclosure regulation to proxy for changes in stock price informativeness (Foucault and Frésard 2012; Edmans, Jayaraman, and Schneemeier 2017; Jayaraman and Wu 2019). Besides learning from the firm's own stock price, managers also have been found to learn additional information from peers' stock prices (Foucault and Frésard 2014; Dessaint et al. 2019; Yan 2017). The managerial learning channel has been shown to play an important role in shaping firms' product market strategy (Foucault and Frésard 2019) and compensation contracts (Lin, Liu, and Sun 2019). Our analysis of the ETF pathway provides further support to this channel.[9]

Before closing the introduction, it is worth considering why ETF ownership, as opposed to inclusion in sector funds, is crucial for the managerial learning

7 The empirical findings of Boehmer and Boehmer (2003) show that the initiation of ETFs increases liquidity and market quality. Li and Zhu (2022) argue that, because of the high liquidity and creation-redemption mechanism, ETFs can relax short-sale constraints for difficult-to-short stocks. Dannhauser (2017) finds that corporate bond ETFs have a long-term positive valuation effect on their constituents.

8 See, for example, Subrahmanyam and Titman (1999), Goldstein, Ozdenoren, and Yuan (2013), and Sockin and Xiong (2015) in the context of equity and commodity markets, respectively.

9 Brogaard, Ringgenberg, and Sovich (2018) provide evidence that the profits of firms with significant exposure to index commodities are adversely affected following the financialization of commodity markets. Our paper examines the cross-sectional effects of ETF ownership on common versus firm-specific information, whereas their paper eschews this topic.

predictions. Note that unlike open-ended funds, ETFs are liquid and tradeable, thus providing factor-informed traders with an instrument devoid of the concern about trading against investors with firm-specific information.[10] In addition, investors can take a leveraged position or short-sell ETFs, further enhancing the attractiveness of these securities for factor-informed traders.

1. The Model

In this section, we provide a simple model that motivates our empirical analysis. The model links real decisions to financial markets. The literature shows that when a firm's stock price affects and reflects investment decisions, the stock price is typically nonlinear (e.g., Goldstein, Ozdenoren, and Yuan 2013; Sockin and Xiong 2015). For simplicity and for the purpose of guiding the empirical analysis, we follow the approach of Dessaint et al. (2019) and Subrahmanyam and Titman (1999), where the firm's investment is in a growth opportunity, whereas the traded security is a claim to its assets in place. That is, the stock price of the established business (i.e., assets in place) influences, and does not reflect, the investment decision of developing a new product/business (i.e., growth opportunity). The fundamentals of the new business are related to those of the established business. We now present and then use the model to derive testable implications. The proofs of all claims appear in Section A.1 in Appendix A.

1.1 Model setup

The payoff on the assets in place is

$$v=\zeta+\beta+\theta, \tag{1}$$

where the three terms on the right-hand side are mutually independent. In Equation (1), we view ζ and β as composite variables that represent common, or systematic, components of firm value (with each related to both macroeconomic and industry/sector information flows). We interpret θ as an idiosyncratic component. The prior distributions of the three components are $\zeta\sim N(\mu_\zeta,\tau_\zeta^{-1})$, $\beta\sim N(\mu_\beta,\tau_\beta^{-1})$, and $\theta\sim N(\mu_\theta,\tau_\theta^{-1})$. For simplicity and without loss of generality, we normalize $\mu_\zeta=\mu_\beta=\mu_\theta=0$ throughout the paper. The claim on the assets in place is traded in a one-period Kyle (1985) set-up, with a liquidity or noise trade in the amount of $e\sim N(0,\tau_e^{-1})$. The standard Kyle (1985) assumptions apply.

The informational structure and real investment in the model are as follows:

1. The numbers of traders with information about the three terms ζ, β, and θ are n_1, n_2, and n_3, respectively.

[10] Li and Zhu (2022) and Huang, O'Hara, and Zhong (2021) provide evidence that ETFs can be used as arbitrage instruments to help improve the efficiency of underlying securities' prices.

2. The signals for the three types of informed traders are $\zeta+\varepsilon_1$, $\beta+\varepsilon_2$, and $\theta+\varepsilon_3$, respectively. For simplicity and without loss of generality, we assume that $\varepsilon_1 \to 0, \varepsilon_2 \to 0$ and $\varepsilon_3 \to 0$; that is, they receive perfect information.
3. The quantity n_1 is exogenous, whereas n_2 and n_3 are determined in equilibrium. The costs of acquiring information about β and θ are given by c_2 and c_3, respectively.
4. At $t=0$, the firm's manager has a real investment project (or a growth opportunity). After making an investment K at $t=0$, the firm realizes the project's payoff at $t=1$ as

$$Y(K)=vK;$$

and the cost of the investment is $\frac{1}{2}K^2$.

We now list and explain the assumptions that guide our analysis.

Assumption 1. The number of traders informed about ζ, n_1, is an increasing function of ETF ownership ω, and the cost of acquiring information about β, c_2, is a decreasing function of n_1.

The logic is that ETF ownership promotes trading on the common factor ζ, and the assumption that the number of traders informed about ζ increases in ω is a reduced-form way of modeling this aspect. We provide further details and justification for this assumption in Section A.2 in Appendix A.[11] We also assume that the greater the number of ζ-informed traders (i.e., the more the analysis devoted to ζ), the cheaper it is to obtain information about β. This is a reduced-form way of modeling the idea that the ease of obtaining information about β is increasing in stock price informativeness about ζ. For example, the more the attention paid to the supply/demand for oil, the easier it is to uncover information about demand for services in, say, the transportation sector. As another example, the more the attention paid to aggregate corporate profits (whether they emanate from revenues or from variable costs), the easier it is to assess cash flows to companies with high variable costs, like supermarkets. As yet a third instance, the more the analysis of consumer spending and its components, the easier it is to ascertain demand for an industry's specific products, like smartphones.[12]

[11] For brevity, we do not endogenize the ETF or its price in the main paper, but provide the details in the Appendix. The argument therein motivates Assumption 1 by appealing to the notion that ETFs stimulate basket-based liquidity or noise trading as in Subrahmanyam (1991). Trades from such agents subsidize factor information collection, which spills over to the underlying securities.

[12] We include β and ζ to capture a direct effect of ETFs as well as complementarities in information acquisition. However, omitting any one of these will not affect the main results in our model.

The manager has private information about asset payoffs. The first managerial signal is about $\zeta+\beta$, that is,

$$\chi=\zeta+\beta+\varepsilon_\chi,$$

where $\varepsilon_\chi \sim N(0,\tau_\chi^{-1})$. The second signal is about θ, that is,

$$s=\theta+\varepsilon_s,$$

where $\varepsilon_s \sim N(0,\tau_s^{-1})$. Both ε_s and ε_χ are independent of each other and of other random variables.[13]

1.2 Equilibrium

Write the demand from each informed trader of the three types as x_j, y_j, and z_j, where the subscript j denotes an individual informed trader j. Let μ_ν and τ_ν, respectively, denote the mean and the precision of a generic random variable ν. The equilibrium in the financial market, characterized by $(\lambda_1,\gamma_1,\eta_1,\kappa_1,n_2,n_3)$, consists of three elements:

(1) The market maker sets a linear pricing rule

$$p\left(\sum_{n_1} x_j+\sum_{n_2} y_j+\sum_{n_3} z_j+e\right)=\lambda_1\left(\sum_{n_1} x_j+\sum_{n_2} y_j+\sum_{n_3} z_j+e\right); \qquad (2)$$

(2) informed traders use symmetric linear trading strategies

$$x_j=x(\zeta)=\gamma_1\zeta,\;\; y_j=y(\beta)=\eta_1\beta,\, z_j=z(\theta)=\kappa_1\theta; \qquad (3)$$

and (3) the competitive market means that the ex ante expected net profit for an informed speculator is zero, that is,

$$\mathbb{E}[\pi_2(\beta)]-c_2=0,\;\; \mathbb{E}[\pi_3(\theta)]-c_3=0. \qquad (4)$$

First, we present the equilibrium with exogenous values of n_1, n_2, and n_3.

Lemma 1. In equilibrium, the pricing rule is given by Equation (2), where

$$\lambda_1=\left[\tau_e\left(\frac{n_1}{(n_1+1)^2}\frac{1}{\tau_\zeta}+\frac{n_2}{(n_2+1)^2}\frac{1}{\tau_\beta}+\frac{n_3}{(n_3+1)^2}\frac{1}{\tau_\theta}\right)\right]^{\frac{1}{2}}, \qquad (5)$$

and the trading strategies of informed traders are given by Equation (3), where $\gamma_1=1/[\lambda_1(n_1+1)]$, $\eta_1=1/[\lambda_1(n_2+1)]$, and $\kappa_1=1/[\lambda_1(n_3+1)]$.

[13] It is possible to model more aspects of ETFs, such as the notion that they might increase the amount of index-based noise trading in the individual stocks or that they might provide an additional (noisy) signal to managers about ζ. We note two points: First, as in Subrahmanyam (1991), increased noise trading subsidizes information collection and thus stimulates informed trading in our model. Thus, increased noise trading would tend to increase price informativeness, so that our results on ETFs and price informativeness would continue to obtain. Second, as long as ETFs provide only a noisy signal via prices to management, our results would survive. Thus, these extensions lead to similar results under a wide parameter range as those we present; details are available from the authors.

Next, we present an equilibrium result under endogenous values of n_2 and n_3, which are pinned down by the additional equilibrium conditions, Equation (4). We have the following lemma.

Lemma 2. Both n_2 and n_3 are increasing in ω.

Our structure implies two forces that drive an increase in n_2, and one force that drives an increase in n_3, which implies that n_2 has a tendency to respond more strongly than n_3 to a change in ω. The intuition is the following. First, when the number of traders informed about one component of cash flow increases, the profit from trading on another component goes up for a given number of informed traders with signals about the other component. This is the competition effect in the Kyle framework (see, e.g., Subrahmanyam 1991). In our model, an increase in n_1, caused by a higher ETF ownership ω, promotes entry and results in an increase in n_2 and n_3. Moreover, an increase in n_2 and an increase in n_3 reinforce each other due to the aforementioned competition mechanism.[14] Second, an increase in n_1, caused by a higher ETF ownership ω, also lowers c_2 (Assumption 1), which further increases n_2, via the indifference condition (4).

1.3 Model implications

Next, we analyze the model's implications for market efficiency and real investment.

1.3.1 Stock price informativeness. As in Brunnermeier (2005) and Goldstein and Yang (2019), stock price informativeness is measured by the inverse of residual uncertainty, i.e., by the reciprocal of $\mathrm{var}(\cdot|p)$. We have Proposition 1.

Proposition 1. As ETF ownership (ω) increases, price informativeness about v increases.

ETF ownership encourages collection of common information as well as firm-specific information (i.e., n_1, n_2, and n_3 increase in ω). As a result, it increases stock price informativeness. In fact, in the proof in the appendix, we show

$$p=\left(\frac{n_1}{n_1+1}\zeta+\frac{n_2}{n_2+1}\beta+\frac{n_3}{n_3+1}\theta\right)+\lambda_1 e,$$

which implies that when n_1, n_2, or n_3 increases, the coefficient in front of the corresponding fundamental factor increases while the coefficient in front of noise trading, λ_1, decreases (by Equation (5)).

14 It is interesting to note that the result—more trading on one factor encourages trading on other factors—is also true in the Grossman-Stiglitz REE framework (see, e.g., Goldstein and Yang 2019; Benhabib, Liu, and Wang 2019), where more trading on one factor increases the price informativeness about that factor and thus reduces the risk of the total fundamental value faced by other types of traders who then would have incentives to trade more on other factors.

1.3.2 Real investment. We now solve for the optimal investment policy. Let $\mathcal{I}$ represent the information set of the manager. Then the firm manager's optimal investment decision is given by

$$K^* = \max_K \mathbb{E}\left[\left(vK - \frac{1}{2}K^2\right)\bigg|\mathcal{I}\right],$$

which implies that $K^* = \mathbb{E}[v|\mathcal{I}]$. It is easy to see that the expected profit from real investment is an increasing function of the precision of $\mathcal{I}$. Since $\mathcal{I} = \{p, \chi, s\}$, we have

$$K^* = \mathbb{E}(v|p, \chi, s) = b_1 p + b_2 \chi + b_3 s, \tag{6}$$

where the expressions for b_1, b_2, and b_3, as functions of (n_1, n_2, n_3), are given in the appendix. When ω increases, n_1, n_2, and n_3 increase, so b_1 unambiguously goes up. Proposition 2 follows.

Proposition 2. The sensitivity of the firm's investment to the price (b_1) increases with ETF ownership (ω).

Because n_1, n_2, and n_3 all increase in ω, we are able to prove analytically the result in Proposition 2. When ETF ownership (ω) increases, prices become more informative about the fundamental value $v = \zeta + \beta + \theta$. The firm manager learns from the market price about fundamentals, so real investment is more sensitive to p when ETF ownership is higher. The firm manager has incentives to learn from the stock price about the firm-specific component θ in addition to the factors ζ and β, which is consistent, for example, with the arguments of Luo (2005) and Chen, Goldstein, and Jiang (2007).

The term b_1 in Equation (6) reflects how stock prices allow managers to learn about fundamentals and thus guide their real investment decisions. The coefficients b_2 and b_3 reflect the channel that directly flows from the manager's signals χ and s to real investment. An interesting implication of our model is that b_2 and b_3 may *decrease* in ETF ownership.[15] This is because the manager relies less on own information and more on prices in making investment decisions as ETF ownership rises (because the price becomes more informative about $\zeta + \beta$ as well as about θ). Thus, ETF ownership may actually reduce the reliance of real investment on the manager's own information, and thus strengthen the learning channel pathway. We demonstrate this phenomenon in Figure 2.

1.4 Model extension

We now assume that the firm's manager also learns from the stock prices of peer firms. These prices provide additional signals about $\zeta + \beta$.[16] For simplicity,

[15] We are able to analytically prove this result under some sufficient condition (see Proposition 3 and its proof).

[16] See Foucault and Frésard (2014) and Dessaint et al. (2019).

and as a reduced form, we assume that the additional signal provided by peer firms' stock prices for the firm manager is

$$\rho = \zeta + \beta + \varepsilon_\rho,$$

where $\varepsilon_\rho \sim N(0, \tau_\rho^{-1})$ is independent of all other random variables. The firm manager still receives noisy private signals χ and s as specified earlier. The manager's information set becomes $\mathcal{I} = \{p, \chi, s, \rho\}$ and the manager's investment decision is hence given by

$$K^* = \mathbb{E}(v \mid p, \chi, s, \rho) = b_1 p + b_2 \chi + b_3 s + b_4 \rho.$$

Proposition 3 follows.

Proposition 3. Under the sufficient condition that τ_θ is high enough such that $\frac{\partial n_2}{\partial \omega} / \frac{\partial n_3}{\partial \omega}$ is not too low, b_4 is decreasing in ω. That is, as ETF ownership of the firm increases, the sensitivity of the firm's investment to peer firms' prices decreases.

As the manager's own signal χ becomes extremely precise (i.e., as $\tau_\chi \to \infty$), the weight on the peer signal b_4 goes to zero. Thus, the peer signal is useful if managers' signals are noisy enough that they learn from both own firms' and peer firms' stock prices. Provided this is the case, when the firm's ownership by ETFs increases, the informativeness of the firm's own stock price about $\zeta +\beta$ increases, so the manager finds own (peers') prices more (less) useful.[17] We test Proposition 3 in Section 4.3.

1.5 Numerical simulation

We provide a numerical example. Note that the main results of our model—Lemma 2 and Propositions 1 and 2—are proved analytically, and the numerical simulation is only necessary for the result in Proposition 3. But to help grasp the overall intuition, we also use this example to illustrate the results in Lemma 2 and Proposition 2. We consider the following parameter set: $\tau_\zeta = 20$, $\tau_\beta = 20$, $\tau_\theta = 20$, $n_1(\omega) = 10 + 70\omega$, $\tau_e = 10$, $c_2(n_1(\omega)) = 0.002 - 0.005\omega$, $c_3 = 0.002$, $\tau_\chi = 40$, $\tau_s = 40$, and $\tau_\rho = 30$.[18] Figure 1 depicts the result in Lemma 2. As can be seen, n_2 increases more steeply than n_3 in response to an increase in ω, so that the direct effect of ETF ownership (to increase n_2) is stronger than the indirect

[17] Proposition 3 is true under certain conditions because ETF ownership increases the stock price informativeness about both factor $\zeta+\beta$ and factor θ. In the rare and uninteresting case in which the increase in price informativeness about $\zeta+\beta$ is much weaker than about θ, the increase of b_1, reflecting the overall increase of price informationveness about $\zeta+\beta$ and θ, represents an "overweighting" on price p regarding factor $\zeta+\beta$ because the price informativeness about $\zeta+\beta$ does not increase much. To "cancel" a part of the "overweighting" of b_1, b_2, and b_4 need to also increase.

[18] The parameter values are chosen for illustrative purposes; we have verified that the results hold for a large parameter space.

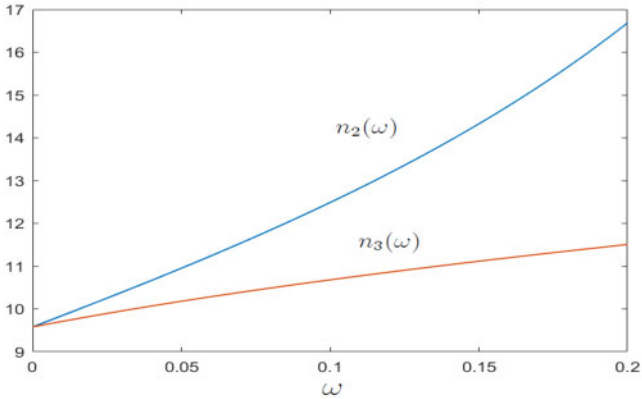

Figure 1
The result within Lemma 2:
This figure depicts the numbers of informed traders n_2 and n_3 as a function of ETF ownership ω. Parameter values are listed in Section 1.5.

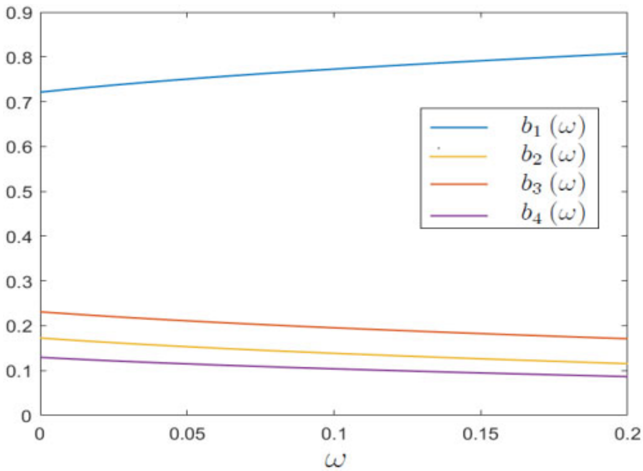

Figure 2
Investment-price sensitivity in response to a change in ω:
This figure depicts the sensitivity of real investment to market price [$b_1(\omega)$] and to the stock prices of peer firms [$b_4(\omega)$] (as well as to the firm manager's private signals [$b_2(\omega)$ and $b_3(\omega)$]), as functions of ETF ownership ω. Parameter values are listed in Section 1.5.

effect (to increase n_3). Thus, the figure indicates that the ETF-induced increase in incentive to conduct informed trading about the common component β is stronger than that about the firm-specific component θ. Figure 2 demonstrates the results in Propositions 2 and 3. As pointed out in the discussion following Proposition 2, the coefficients b_2 and b_3 also decrease in ω.

1.6 Additional results

In this section, we provide additional cross-sectional implications of the model.

1.6.1 The precision of β. With a higher precision of β, τ_β, the informational advantage of β-informed traders decreases, and so expected profits to these traders decrease. This, in turn, reduces the incentive for traders to acquire information about β. In the extreme case when τ_β is very high, few traders might wish to acquire information about β even if the information cost is close to zero. In other words, when τ_β is very high, n_2, and, in turn, b_1, are relatively insensitive to ω. These observations imply the following analytical prediction:

Cross-Sectional Prediction 1. $\partial b_1\left(\omega;\tau_\beta=\tau_\beta^L\right)/\partial\omega>\partial b_1\left(\omega;\tau_\beta=\tau_\beta^H\right)/\partial\omega$ for some $\tau_\beta^H>\tau_\beta^L$. That is, when τ_β is higher, the positive effect of ETF ownership on investment-stock price sensitivity is weaker.

Note that τ_β may be higher for two reasons: first, the inherent volatility of the common factor may be higher, and second, the sensitivity of the firm's return to the factor may be larger in absolute terms.

1.6.2 The manager's signal precision. Recall that the firm's manager has private signals: $s=\theta+\varepsilon_s$ where $\varepsilon_s\sim N(0,\tau_s^{-1})$, and $\chi=\zeta+\beta+\varepsilon_\chi$ where $\varepsilon_\chi\sim N(0,\tau_\chi^{-1})$. We now allow the manager to trade on private information, and investigate the effect of the manager's signal precision τ_s on real investment as well as on trading profits.[19]

To obtain intuition, we first consider two extremes of the signal precision τ_s, namely, $\tau_s=0$ and $\tau_s=+\infty$. For simplicity and convenience, we then have the following three cases for the two sets of precision $\left(\tau_s,\tau_\chi\right)$.

Case 1: $\left(\tau_s=+\infty,\tau_\chi=0\right)$ and $\left(\tau_s=0,\tau_\chi=0\right)$. For $\left(\tau_s=+\infty,\tau_\chi=0\right)$, we show in Appendix A that $\partial b_1(\omega)/\partial\omega>0$ as in Proposition 2, and that the expected trading profit for the manager is positive. For $\left(\tau_s=0,\tau_\chi=0\right)$, the manager does not have any private information, i.e., the manager's information set is the same as the market maker's, which implies $\mathbb{E}(v|p)=p$, that is, $b_1=1$ or $\partial b_1(\omega)/\partial\omega=0$; moreover, the expected profit from trading is zero.

Case 2: $\left(\tau_s=+\infty,\tau_\chi\right)$ and $\left(\tau_s=0,\tau_\chi\right)$, where τ_χ is positive and small. Per the result in Case 1, together with Proposition 2, it follows that $\partial b_1(\omega)/\partial\omega$ is higher under $\left(\tau_s=+\infty,\tau_\chi\right)$ than under $\left(\tau_s=0,\tau_\chi\right)$. Moreover, the expected profit from managerial trading is higher under $\left(\tau_s=+\infty,\tau_\chi\right)$ than under $\left(\tau_s=0,\tau_\chi\right)$.

Case 3: $\left(\tau_s=+\infty,\tau_\chi=\tau_\chi^H\right)$ and $\left(\tau_s=0,\tau_\chi=\tau_\chi^L\right)$, where τ_χ^H and τ_χ^L are positive and small and $\tau_\chi^H-\tau_\chi^L\geq 0$ is not large. As long as $\tau_\chi^H-\tau_\chi^L\geq 0$ is not large, the result in Case 2 carries over.

[19] We focus on τ_s, rather than τ_χ, as we empirically proxy for signal precision via the profitability of insider trading (Section 4.5), which is more likely to reflect firm-specific signals. Nonetheless, similar results apply for τ_χ.

Write π_I as the expected profit to the manager from trading. We then have our second prediction:

Cross-Sectional Prediction 2. $\partial b_1(\omega;\tau_s=\tau_s^H,\tau_\chi=\tau_\chi^H)/\partial\omega > \partial b_1(\omega;\tau_s=\tau_s^L,\tau_\chi=\tau_\chi^L)/\partial\omega$ and $\pi_I\left(\tau_s=\tau_s^H,\tau_\chi=\tau_\chi^H\right) > \pi_I\left(\tau_s=\tau_s^L,\tau_\chi=\tau_\chi^L\right)$ for some $\tau_s^H>\tau_s^L$ when τ_χ^H and τ_χ^L are small enough and $\tau_\chi^H-\tau_\chi^L\geq 0$ is not too large. That is, for greater τ_s (while τ_χ is held fixed at a sufficiently low level, or allowed to increase by a sufficiently small amount from that level), the positive effect of ETF ownership on investment-stock price sensitivity is stronger and the expected profit from managerial trading is higher.

The intuition is that when firm managers have very imprecise firm-specific information, they put virtually all their Bayesian weight on the market price, so that investment sensitivity to market price, which depends directly on this weight, is virtually at its maximum and does not shift much with ETF ownership. If managers have high-quality firm-specific information, however, the sensitivity is very responsive to the additional factor information introduced by ETF ownership.

In the ensuing empirical analysis, we test Proposition 2, Propositions 1 and 3, and the two additional predictions listed above. Given the observations in Section 1.5, we focus on the direct effect of ETF ownership in increasing flows of common information.[20] After presenting our data (Section 2), we test Proposition 2 within Section 3. Since market index prices are readily available, we then empirically investigate the notion that nonmarket ETF information is more likely to be useful to the firm manager (Section 4.1).[21] Subsequently, we provide tests of Propositions 1 and 3, and the cross-sectional implications listed above.

2. Data

In this section, we describe the data sources and the calculation of the key variables used in the empirical analysis. We obtain a list of all U.S. domestic equity ETFs that physically replicate the indexes.[22] We do so by first merging all ETFs (where the variable fet_flag = F) in the CRSP mutual fund database with securities in the CRSP monthly stock file with the share code of 73. We then

20 See, however, the discussion in Section 4.3.

21 In our model, ETFs induce an increase in the number of factor-informed traders, which stimulates firm-specific information acquisition. The model, however, can accommodate a different specification in which the total number of potentially informed agents is fixed. In this specification, a substitution effect is also possible, wherein ETFs prompt traders to focus on factor information to the *exclusion* of firm-specific information. Such a variation leads to similar results as long as the complementarity dominates (details are available on request). Our modeling choice is consistent with most of the recent empirical literature (Glosten, Nallareddy, and Zou 2021; Bhojraj, Mohanram, and Zhang 2020; Huang, O'Hara, and Zhong 2021), which indicates ETFs increase price informativeness. Our evidence to follow (see Tables 3 and 6) is consistent with this approach as well.

22 Most ETFs in the U.S. tend to replicate their underlying index. The Investment Act of 1940 requires ETFs to hold 80% of their assets in securities matching the fund's name.

parse fund names to tease out nonequity or nondomestic ETFs.[23] Finally, we require that the ETFs have holdings information available from the Thomson Reuters Mutual Fund holdings database (S12). Our final sample contains 605 ETFs from 2003 to 2016. We construct ETF ownership (ETF_{it}) of each stock i in year t using the following equation:

$$ETF_{it} = \frac{\sum_{j=1}^{J} SHARES_{ijt}}{TSO_{it}}$$

where $SHARES_{ijt}$ is the number of shares of firm i held by ETF j at the end of year t and TSO_{it} is firm i's total number of shares outstanding at the end of year t.

We obtain stock price and return information from CRSP, and accounting data from Compustat. We restrict the sample to stocks traded on NYSE, AMEX, and Nasdaq, and exclude financial and utility firms. We further exclude observations without necessary data (investment and standard control variables), and filter out observations with sales and asset growth larger than 100% and total assets less than $1 million. Glosten, Nallareddy, and Zou (2021) find that the effect of ETF ownership on the informativeness of stock prices is much greater for midcap and small companies, which tend to have more opaque information environments. Motivated by this finding, and since our focus is on firms most likely to benefit from ETF ownership, we exclude from our sample those firms whose market capitalization is ranked in the top 20% of the distribution each year. We winsorize all variables at the 1% and 99% levels each year to mitigate the potential effect of outliers.

Table 1 reports the summary statistics of the variables used in our analysis. The average ETF ownership in our sample is 4.4% with a standard deviation of 3.5%. Figure 3, which plots the time series of average ETF ownership, indicates that it has risen over time, from less than 1% in 2000 to around 8% in 2016. The means (standard deviations) of the investment variables, *CAPXRND*, *CAPX*, and *RND* are 0.108 (0.116), 0.051 (0.062) and 0.056 (0.103), respectively. This indicates that a firm's annual investment represents about 10.8% of its total assets, and is attributed almost equally to capital expenditures and R&D expenses. The mean q is 1.93, close to what is typically reported in the literature (e.g., Chen, Goldstein, and Jiang 2007). Panel B of Table 1 shows statistics related to the number, size and holdings of market and nonmarket ETFs.

3. The Basic Empirical Results

In this section we present our main empirical findings, which involve tests of Proposition 2. First, we present our baseline results, along with various

[23] Specifically, we drop funds with the following words in the variable *lipper class name* in the CRSP files: "International", "Global", "World", "Japan", "Japanese", "European", "Emerging Markets", "China", "India", "Latin", "World", "Pacific", "Leverage", "Short Bias", "Alternative", "Mixed Asset", "Gold", "Natural Resources" or "Real Estates" .

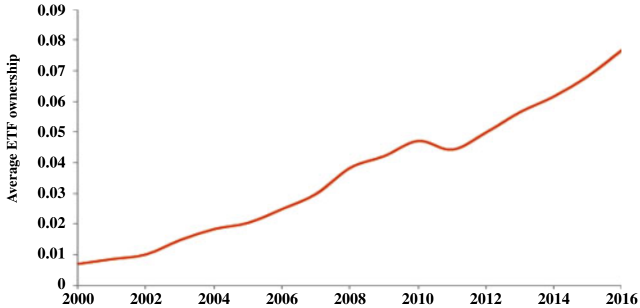

Figure 3
Average ETF ownership by year:
This figure plots the average fraction of shares outstanding held by ETFs for firms in our sample from years 2000 to 2016. The vertical axis indicates the magnitude of ETF ownership, and the horizontal axis indicates the year. Our method for calculating ETF ownership is described in Section 2.

robustness tests. We then discuss an instrumental variable approach to address endogeneity concerns, and finally, we examine if ETF ownership raises stock price informativeness with respect to common information (Proposition 1).

3.1 ETF ownership and investment-price sensitivity

To examine whether ETF ownership (ETF_{it-1}) affects the sensitivity of a firm's investment ($Investment_{it}$) to its own (normalized) stock price (Q_{it-1}) (as in Proposition 2), we estimate the following regression:

$$Investment_{it} = \alpha_t + \chi_i + d_1 Q_{it-1} + d_2 Q_{it-1} * ETF_{it-1} + d_3 ETF_{it-1} + \psi X_{it-1} + \epsilon_{it}, \quad (7)$$

$Investment_{it}$ is firm i's investment in year t measured by the sum of capital expenditures and R&D expenses, capital expenditures, and R&D expenses, all scaled by lagged total assets. Q_{it-1} is firm i's Tobin's q in year $t-1$, defined as the market value of equity plus the book value of assets minus the book value of equity, scaled by the book value of assets at the end of the previous year. The key right-hand variable is the interaction term $Q_{it-1} * ETF_{it-1}$, which captures the incremental effect of ETF ownership on investment-q sensitivity.

Our specification controls for various firm characteristics (X in Equation (7)) known to affect investments and their sensitivity to stock prices. To account for the positive effect of cash flows on investments (Fazzari, Hubbard, and Petersen 1988), we include cash flow, both on its own and as an interaction with ETF (CF_{it} and $CF_{it} * ETF_{it-1}$). Since the size of a firm may correlate with its investment opportunities (e.g., Bushee 1998; Foucault and Frésard 2014; Jayaraman and Wu 2019), we control for size, both on its own and its interaction with Tobin's q ($SIZE_{it-1}$ and $SIZE_{it-1} * Q_{it-1}$). To ensure that the ETF variable is not a proxy for institutional ownership in general, we control for the latter, which we measure similarly to ETF ownership. Given the high correlation between institutional and ETF ownership, we follow

Table 1
Descriptive statistics

A. Summary statistics of main variables

Variables	N	Mean	SD	P25	Median	P75
$CAPXRND_{it}$	22,524	0.108	0.116	0.033	0.070	0.140
$CAPX_{it}$	22,524	0.051	0.062	0.015	0.030	0.060
RND_{it}	22,524	0.056	0.103	0.000	0.005	0.073
Q_{it-1}	22,524	1.926	1.302	1.130	1.498	2.203
PQ_{jt-1}	20,722	2.124	0.848	1.473	1.927	2.581
ETF_{it-1}	22,524	0.044	0.035	0.016	0.036	0.066
$NonMktETF_{it-1}$	22,524	0.031	0.029	0.005	0.024	0.048
$MktETF_{it-1}$	22,524	0.013	0.009	0.006	0.012	0.019
ROA_{it+1}	21,807	0.076	0.230	0.040	0.109	0.170
SG_{it+1}	21,762	0.391	25.872	−0.038	0.058	0.166
CF_{it}	22,524	0.037	0.179	0.009	0.075	0.125
$SIZE_{it-1}$	22,524	5.959	1.378	5.025	6.086	7.035
$INST_{it-1}$	22,524	0.616	0.280	0.399	0.669	0.849
$INSTR_{it-1}$	22,524	0.013	0.192	−0.124	0.018	0.149
RET_{it+3}	22,524	0.024	0.319	−0.157	0.033	0.200
SG_{it-1}	22,524	0.124	0.362	−0.026	0.072	0.197
$CASH_{it-1}$	22,524	0.225	0.227	0.045	0.143	0.336
LEV_{it-1}	22,524	0.176	0.186	0.001	0.126	0.294
$1/ASSET_{it-1}$	22,524	0.008	0.014	0.001	0.003	0.008
$ETFNum_{it-1}$	22,524	2.650	0.947	1.946	2.890	3.401
$\Delta ASSET_{it}$	22,524	0.059	0.213	−0.049	0.039	0.144
$M\&A_{it}$	22,524	0.030	0.077	0.000	0.000	0.014
$CAPXRND_{iq}$	91,230	0.027	0.031	0.007	0.017	0.035
$CAPX_{iq}$	91,180	0.012	0.016	0.003	0.007	0.014
RND_{iq}	91253	0.014	0.028	0.000	0.000	0.019
Q_{iq-1}	91,253	1.951	1.347	1.130	1.506	2.233
ETF_{iq-1}	91,253	0.043	0.034	0.015	0.035	0.064

B. Statistics for nonmarket and market ETFs

	Nonmarket ETFs	Market ETFs
# of unique ETFs	531	74
Mean mktcap (millions $)	517	7,088
Mean # of stocks held	186	620

Panel A reports summary statistics for the main variables used in the analysis (see Appendix B for variable definitions). Panel B reports the number of unique ETFs, the average market capitalization, and the average number of stocks held by nonmarket and market ETFs separately. Market ETFs include those ETFs that physically track broad market indexes, including S&P 500, S&P 1500, Russell 1000, Russell 3000, and the NYSE/ NASDAQ Composite Index. Nonmarket ETF ownership is defined as firm ownership by ETFs not classified as market ETFs.

Glosten, Nallareddy, and Zou (2021) and use residual institutional ownership $INSTR_{it-1}$, after orthogonalizing it to ETF ownership. We also include the interaction between $INSTR_{it-1}$ and Tobin's q because institutional ownership can affect the informational efficiency of stock prices (Boehmer and Kelley 2009) and hence the investment-q sensitivity.

To address the tendency of overvalued firms to invest more (Baker, Stein, and Wurgler 2003; Polk and Sapienza 2009), we control for annualized firm stock returns over the next 3 years (RET_{it+3}).[24] To account for investment constraints and operating performance, our model includes leverage (LEV_{it-1}), cash holdings ($CASH_{it-1}$), return on assets (ROA_{it-1}), and sales growth

[24] We require a stock to have at least 1 year of future returns to construct this variable.

Table 2
ETF ownership and investment-*q* sensitivity

	$CAPXRND_{it}$	$CAPX_{it}$	RND_{it}
$Q_{it-1} \times ETF_{it-1}$	0.145***	0.037***	0.109***
	(5.09)	(2.73)	(5.00)
Q_{it-1}	0.053***	0.010***	0.041***
	(8.25)	(3.71)	(7.78)
ETF_{it-1}	−0.420***	−0.118***	−0.297***
	(−6.20)	(−2.96)	(−6.33)
CF_{it}	−0.046***	0.042***	−0.085***
	(−3.50)	(8.31)	(−7.81)
$CF_{it} \times ETF_{it-1}$	0.191	−0.039	0.107
	(0.73)	(−0.44)	(0.49)
$SIZE_{it-1}$	0.004**	0.006***	−0.002**
	(2.11)	(4.99)	(−1.97)
$SIZE_{it-1} \times Q_{it-1}$	−0.006***	−0.001**	−0.005***
	(−6.29)	(−2.20)	(−6.41)
$INSTR_{it-1}$	−0.029***	−0.010*	−0.022***
	(−2.96)	(−1.79)	(−3.40)
$INSTR_{it-1} * Q_{it-1}$	0.012**	0.005**	0.008**
	(2.56)	(2.07)	(2.46)
RET_{it+3}	−0.009***	−0.003*	−0.007***
	(−3.65)	(−1.74)	(−3.80)
SG_{it-1}	0.005*	0.003***	0.002
	(1.88)	(2.97)	(0.80)
$CASH_{it-1}$	−0.000	−0.005	0.005
	(−0.07)	(−1.27)	(0.83)
LEV_{it-1}	−0.076***	−0.040***	−0.031***
	(−9.71)	(−9.33)	(−5.50)
ROA_{it-1}	−0.035***	0.008***	−0.038***
	(−4.99)	(3.06)	(−6.74)
$1/ASSET_{it-1}$	1.126***	0.281***	0.860***
	(5.30)	(3.39)	(5.21)
Adjusted R^2	.789	.689	.900
Fixed effects	Y, F	Y, F	Y, F
No. of obs.	21922	21922	21922

This table presents the results from the regression of firm investments ($CAPXRND_{it}$, $CAPX_{it}$ and RND_{it}) on the interaction of Tobin's q and ETF ownership ($Q_{it-1} * ETF_{it-1}$). We exclude from our sample those firms whose market capitalization ranked in the top 20% of the distribution. Both firm and year fixed effects are included. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. See Appendix B for variable definitions. $*p<.1$; $**p<.05$; $***p<.01$.

(SG_{it-1}) (Foucault and Frésard 2012; Panousi and Papanikolaou 2012). Since investments and Q are scaled by total assets, we control for the inverse of total assets ($1/ASSET_{it-1}$) to ensure that our findings are not driven by the common deflator (Chen, Goldstein, and Jiang 2007). Finally, to account for any unobserved time-invariant firm-specific factors and variation in investments over time, all our models include firm and time fixed effects, denoted by χ_i and α_t, respectively, in Equation (7). The standard errors are clustered at the firm level. This econometric specification is common in empirical corporate finance studies (e.g., Chen, Goldstein, and Jiang 2007; Peters and Taylor 2017). Appendix B provides detailed definitions for the variables used in the analysis.

Table 2, presents our baseline regression results from estimating Equation (7). Consistent with prior studies, a firm's investment shows a significant positive relation with its own stock price for all three measures of investment. Column

1 shows that for a firm not held by any ETFs, a one-standard-deviation increase in Tobin's q leads to an increase of about 4.3 percentage points in a firm's investment, as measured by the sum of capital expenditures and R&D expenses. Columns 2 and 3 show that both capital expenditures and R&D expenses are similarly positively related to Tobin's q.

In line with the Proposition 2, the coefficient for the interaction $Q_{it-1} * ETF_{it-1}$ is positive and significant for all three investment measures, with a magnitude of 0.145 (t=5.09) for $CAPXRND_{it}$, 0.037 (t=2.73) for $CAPX_{it}$, and 0.109 (t=5.00) for RND_{it}. In terms of economic significance, our estimates imply that a one-standard-deviation increase in Tobin's q (1.30) is associated with an increase of 7.2 (8.1) percentage points in corporate investment among firms in the bottom (top) quartile of ETF ownership. This relative increase in investment is economically significant, representing a change of 8.3% relative to average investments in our sample. We also note that the overall effect of ETF ownership on real investment is negative with a magnitude of -0.141 ($d_3 + d_2 * AverageQ_{it-1} = -0.141$). Moreover, the coefficient for the interaction between cash flow and ETF ownership ($CF_{it} * ETF_{it-1}$) is insignificant for all three investment measures, which suggests that the reliance on cash flows for information is not affected by ETF ownership.

In terms of the remaining control variables, we find that firms with higher sales growth and less leverage invest more, indicating that tighter financial constraints and worse operating performance curb corporate investments. The inverse of total assets also has a positive effect on investments, indicating that firms with less assets have greater capacity to grow (Foucault and Frésard 2012, 2014). Institutional ownership exerts a negative effect on corporate investment. This finding has two interpretations: Institutions may encourage managers to pursue short-run performance objectives (Bushee 1998), or they may act as a governance mechanism, curbing managers' tendency to overinvest (Ferreira and Matos 2008). We also find that firms with lower returns in the next 3 years invest more, which suggests a positive relationship between investments and overvaluation (Baker, Stein, and Wurgler 2003; Polk and Sapienza 2009). We find mixed results for cash flow, size, and return on assets across the three different investment measures.[25]

[25] We conduct several sensitivity analyses and confirm that our results continue to hold within Table IA.1 of the Internet Appendix. First, in panel A, we use the number of ETFs holding the stock (*ETFNum*) as an alternative measure of ETF ownership. In panel B, we conduct the analysis at the quarterly frequency. In panel C, we replace residual institutional ownership with raw institutional ownership minus ETF ownership. In panel D, we use alternative measures of investment including the percentage change of total assets, as in Chen, Goldstein, and Jiang (2007), and the amount of money spent on mergers and acquisitions, both on its own and when added to $CAPXRND$. Panel E presents the results that cluster standard errors at both firm and year levels; panel F includes the interaction of Tobin's q with linear and quadratic time trends; and panel G reports results from replacing Tobin's q with Peters and Taylor's (2017) total q (which accounts for intangible capital).

3.2 Instrumental variable model

In this section, to substantiate that endogeneity is not driving our results, we conduct instrumental variable analyses. There are two endogeneity concerns with our previous test. First, it is possible that our specification omits variables that affect the firm's investment-q sensitivity, which may correlate with ETF ownership. A second concern is that stocks with better real investment policies are more likely to be included in ETFs, that is, reverse causality. Even though our models control for a large number of variables, such issues cannot be completely ruled out.

The instrument we use to identify exogenous variation in ETF ownership is proposed by Zou (2019), and is based on the acquisition of Barclays Global Investors (BGI) and its iShares unit by BlackRock at the end of 2009. At that time, Barclays wanted to avoid a possible bailout by the U.K. government, and sold BGI to strengthen its position. Because BlackRock was in a better position to attract capital into its funds, due to a stronger brand name, a more specialized workforce, and better distribution channels (Zou 2019), the assets under management for iShares ETFs increased by 19% one year after the acquisition (BlackRock 2010). This event suggests that stocks with higher iShares ETF ownership (before the acquisition event) should have experienced an exogenous increase in ETF ownership since 2010 relative to those stocks with lower iShares ETF ownership.

We first verify the assumption that Blackrock's acquisition of iShares implies elevated flows to iShares ETFs relative to other ETFs during the post-acquisition period. To that end, we regress monthly ETF flows (as percentage of lagged ETF total net assets) on lagged log of ETF size, past 12-month ETF returns, monthly return volatility of the ETF, and a time trend. We then take the regression residual plus the intercept as the residual flows, and compute the average residual flows to iShares and non-iShares ETFs separately. Figure IA.1 of the Internet Appendix plots the average annual residual flows to iShares and non-iShares ETFs over the 2007–2012 period, conditional on the ETFs existing before the acquisition. We also plot 95% confidence bands around the mean annual residual flows. The figure demonstrates that parallel trends in residual flows to iShares and non-iShares ETFs cannot be rejected prior to the acquisition. However, in the post-acquisition year, iShares ETFs on average experience significantly greater residual flows relative to non-iShares ETFs.

Stocks with high iShares ETF ownership might differ from those with low such ownership along various dimensions. To rule out the possibility that our results reflect these differences, we use a propensity score matching (PSM) method to create a matched sample for stocks with high iShares ownership. The procedure is as follows: First, we use firms' iShares ownership before the acquisition (i.e., year 2009) to define treatment and control groups, setting $Treat$ as one if the firm's iShares ownership is above the sample median, and zero otherwise.[26] We then use a logit model to estimate the probability that a firm is placed in the treatment group from various firm-level characteristics, and match each treated firm to a control firm in the same industry based on the predicted value from the logit model, using the nearest neighbor matching method.[27]

Our instrument is $Post_t * Treat_i$, where $Post_t$ is a dummy that equals one for ETF ownership (ETF_{it}) measured in the years 2010–2013, and zero for 2007–2009.[28] The exclusion restriction is likely to be satisfied because the acquisition was unlikely to have been driven by any fundamental characteristics of the stocks with a larger fraction of shares held by iShares ETFs.

The first-stage regression models are shown below:

$$ETF_{it} = \alpha_t + \chi_i + d_1 Post_t * Treat_i + d_2 Post_t * Treat_i * Q_{it} + d_3 Q_{it} + \psi X_{it} + \epsilon_{it}, \tag{8}$$

$$ETF_{it} * Q_{it} = \alpha_t + \chi_i + d_1 Post_t * Treat_i + d_2 Post_t * Treat_i * Q_{it} + d_3 Q_{it} + \psi X_{it} + \epsilon_{it}, \tag{9}$$

We estimate the above models with both firm and year fixed effects. The sample period spans the years from 2007 to 2013, which is a 7-year period symmetrically centered on the acquisition year. To the extent that treatment and control firms face different financial constraints, their investments may also respond to Tobin's q differently in the period following the financial crisis. Therefore, we control for the interaction of Tobin's q with the text-based financial constraint measure from Hoberg and Maksimovic (2015).[29]

Table 3, panel A, reports the results. The findings in columns 1 and 2 show that treatment stocks experience a significant increase in ETF ownership relative to the control stocks after the acquisition. The F-statistics for both the baseline and interaction instruments are greater than 10, suggesting that the instruments are not weak. In the second stage, we use predicted ETF ownership ($ETF_{it}(IV)$) and predicted $Q_{it} * ETF_{it}$ ($Q_{it} * ETF_{it}(IV)$) from the first-stage regression to reexamine the effect of ETF ownership on investment-q sensitivity as follows:

$$\begin{aligned} Investment_{it} = {} & \alpha_t + \chi_i + d_1 Q_{it-1} + d_2 Q_{it-1} * ETF_{it-1}(IV) + d_3 ETF_{it-1}(IV) \\ & + \psi X_{it-1} + \epsilon_{it}. \end{aligned} \tag{10}$$

Columns 3–5 in panel A of Table 3 report the second-stage regression results. The coefficient for the instrumented interaction is significantly positive

---

26 We measure each stock's iShares ETF ownership using only the iShares ETFs existing before 2009, to address the concern that the increased ETF ownership for treated firms is due to new ETFs launched by BlackRock after the acquisition.

27 As shown in Table IA.2, panel 1, before matching, the differences between treated and control firms are statistically significant for four of seven of the firm characteristics we consider. After the PSM matching, as shown in panel 2 of Table IA.2, these differences are statistically insignificant in all cases, which indicates that the method successfully creates a control group of firms that is similar to the treatment group.

28 Because the acquisition happened at the end of 2009, we consider the acquisition year to be 2010.

29 We thank Jerry Hoberg and Max Maksimovic for making their data available on Hoberg's website.

Table 3
Instrumental variable regressions using BlackRock's acquisition of iShares

	First-stage		Second-stage		
	A. Investment to price sensitivity				
	ETF_{it-1} (1)	$Q_{it-1}\times ETF_{it-1}$ (2)	$CAPXRND_{it}$ (3)	$CAPX_{it}$ (4)	RND_{it} (5)
Post × *Treat*	0.004***	−0.026***			
	(3.04)	(−4.14)			
$Q_{it-1}\times$ *Post* × *Treat*	−0.001	0.018***			
	(−0.94)	(4.59)			
$Q_{it-1}\times ETF_{it-1}(IV)$			0.246***	0.084*	0.175**
			(2.76)	(1.83)	(2.36)
$ETF\ (IV)_{it-1}$			−2.524**	−0.895	−1.480***
			(−2.56)	(−0.99)	(−3.20)
Q_{it-1}	−0.003	−0.075***	0.050***	0.020**	0.028***
	(−1.23)	(−5.31)	(4.88)	(2.50)	(3.52)
Controls	Included	Included	Included	Included	Included
F-test	10.06	27.47			
Adjusted R^2	.845	.887	.791	.740	.927
Fixed effects	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	3,270	3,270	3,270	3,270	3,270
	B. Predicting earnings from stock returns				
	ETF_{it-1} (1)	$RET_{it-1}\times ETF_{it-1}$ (2)	$Earn_{it}$ (3)	$Earn\_Com_{it}$ (4)	$Earn\_Firm_{it}$ (5)
Post × *Treat*	0.008***	−0.002			
	(6.45)	(−0.79)			
$RET_{it-1}\times$*Post* ×*Treat*	−0.000	0.014***			
	(−0.25)	(3.19)			
$RET_{it-1}\times ETF_{it-1}(IV)$			0.936***	0.991***	0.040
			(3.53)	(3.46)	(0.32)
$ETF(IV)_{it-1}$			−0.939	−1.478	0.377
			(−0.49)	(−0.63)	(−0.34)
RET_{it-1}	−0.001***	0.041***	−0.012*	−0.011	−0.006**
	(−3.31)	(5.92)	(−1.84)	(−1.60)	(−2.11)
Controls	Included	Included	Included	Included	Included
F-test	41.67	10.8			
Adjusted R^2	.828	.742	.333	.301	.593
Fixed effects	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	2,778	2,778	2,778	2,778	2,778

This table reports results from instrumental variable (IV) regression analysis, using Blackrock's acquisition of iShares at the end of 2009 to define the relevant instrument. We use stocks' iShares ETF ownership before the acquisition (year 2009) to define treatment and control groups, where iShares ETF ownership is calculated using only the iShares ETFs available before 2009. We define the treatment group as stocks with iShares ETF ownership above the sample median. To conduct the PSM, we define a dummy equal to one for each treated firm and zero otherwise. We then estimate a logit model with this dummy as the dependent variable, and various firm characteristics that may affect corporate investments as regressors. The fitted value from this model is the probability that a firm is placed in the treated sample. We then use this probability to match each treated firm to a control firm within the same industry, using the one-to-one nearest-neighbor matching method. The table reports the IV regression results in the PSM matched sample. *Post* is a dummy variable that is equal to one if ETF_{it-1} is in 2010–2013, and zero if ETF_{it-1} is in 2007–2009. *Treat* is a dummy variable that equals one for stocks with iShares ETF ownership (measured in 2009) above the sample median, and zero for matched control stocks. Columns 1 and 2 report the results from the first-stage regressions, and columns 3–5 report the results from the second-stage regressions, which use the fitted values from the first-stage models. In panel A, we use the IV model to estimate our baseline investment to price sensitivity regression. The dependent variables in the first stage are ETF ownership (ETF_{it-1}) and its interaction with Tobin's q ($Q_{it-1}\times ETF_{it}$). In the second-stage regression, we estimate the model in Equation (8), where the dependent variables are the three investment measures. In panel B, we use the IV model to examine whether stock returns can predict future earnings. Columns 1 and 2 report the first-stage regression results, where the dependent variable is ETF ownership (ETF_{it-1}) and its interaction with annual stock returns ($RET_{it-1}\times ETF_{it-1}$). *Post* and *Treat* are defined as in panel A. Columns 3–5 report the second-stage regression results. In column 3, the dependent variable is firm i's earnings innovations ($Earn_{it}$), and in columns 4 and 5, we decompose $Earn_{it}$ into common and firm-specific components ($Earn\_Com_{it}$ and $Earn\_Firm_{it}$ respectively), which we use as the dependent variables. These variables are regressed on the instrumented interaction term of lagged annual stock return and ETF ownership ($RET_{it-1}\times ETF_{it-1}$) and various lagged controls, including firm size, leverage, return on assets, market-to-book ratio, and institutional ownership orthogonalized to ETF ownership. *Post* and *Treat* are not included in the models individually as they are subsumed by firm and time fixed effects, respectively. See Appendix B for variable definitions. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. *$p<.1$; **$p<.05$; ***$p<.01$.

across all three investment measures, consistent with our hypothesis that ETF ownership facilitates managerial learning from stock prices.[30]

Next, we conduct a test to investigate the type of information that ETFs convey (viz., Proposition 1 of our model). To conduct the test, we regress firm-level changes in earnings at year t ($Earn_{it}$) on the past-year stock return (RET_{it-1}), and its interaction with ETF ownership ($RET_{it-1} * ETF_{it-1}$). To address endogeneity concerns, we apply the IV framework to this test, as in Table 3, panel A, while replacing Q_{it-1} with RET_{it-1}. Columns 1 and 2 of panel B of Table 3 present the first-stage regression results. We find that the coefficients for $Post * Treat$ and $RET_{it-1} * Post * Treat$ are significantly positive and that the F-statistics are greater than 10, suggesting the relevance of the instruments. Column 3 shows that in the second-stage regression, the coefficient for the predicted $RET_{it-1} * ETF_{it-1}$ ($RET_{it-1} * ETF_{it-1}(IV)$) is significant and positive, suggesting that ETFs lead to greater informativeness of stock prices about future earnings.

In columns 4 and 5, we decompose firms' changes in earnings into common ($Earn\_Com_{it}$) and firm-specific components ($Earn\_Firm_{it}$). The method used for this decomposition follows Bhojraj, Mohanram, and Zhang (2020) and is described in Appendix B. In this method, the cross-sectional variation in $Earn\_Com_{it}$ arises from the industry-related component of earnings. We find that the coefficient for $RET_{it-1} * ETF_{it-1}(IV)$ is significantly positive for the common earnings component, but insignificant for the firm-specific one. This result suggests that ETFs facilitate the incorporation of industry-related information into stock prices. It stands to reason that nonmarket ETFs (i.e., those not related to broad market indexes) should be the primary vehicles that transmit such information, and we investigate whether this is the case in Section 4.

In Table IA.5 in the Internet Appendix, we examine whether stock prices are more informative about deeper sources of earnings, specifically, revenues and gross profits, when ETF ownership is high. One motivation for this exercise is that, as we propose in Section 1.1, ETFs can convey information about common factors in product demand. We find that indeed, the informativeness of returns about these sources rises with increasing ETF ownership.

30 We validate the parallel trends assumption for the first-stage difference-in-differences by using the approach in Chen, Kelly, and Wu (2020). Specifically, we examine the dynamic effects of the instrumented variables (ETF ownership and its interaction with q) around the years BlackRock acquired iShares. Thus, we reestimate the models in Equations (8) and (9) by replacing the dummy $Post$ with a series of dummies that flag the years around the acquisition event. If indeed the variables of interest exhibit parallel trends, then we should find the instruments used in each model ($Treat * Post$ in (8) and $Q * Treat * Post$ in (9)) to be statistically insignificant when the time dummy flags a year before the acquisition. The results in Table IA.3 show that this is indeed the case. The parallel trend is visually depicted in Figure IA.2. In Table IA.4, we conduct a similar dynamic analysis of investment-q sensitivity around BlackRock's acquisition of iShares. Again, the results do not show a significant pre-trend, as the coefficients for the triple interactions between q, the treatment dummy, and the time dummies that flag the years before the acquisition are statistically insignificant for all three investment measures.

Table 4
ETF ownership and investment-*q* sensitivity by nonmarket versus market ETFs

	$CAPXRND_{it}$ (1)	$CAPX_{it}$ (2)	RND_{it} (3)
$Q_{it-1} \times NonMktETF_{it-1}$	0.133***	0.050**	0.099***
	(2.98)	(2.55)	(2.82)
$Q_{it-1} \times MktETF_{it-1}$	0.157	−0.025	0.129
	(1.32)	(−0.47)	(1.36)
Q_{it-1}	0.053***	0.011***	0.040***
	(8.04)	(3.80)	(7.66)
$NonMktETF_{it-1}$	−0.472***	−0.173***	−0.315***
	(−5.94)	(−3.62)	(−5.70)
$MktETF_{it-1}$	0.241	0.268*	0.043
	(0.99)	(1.93)	(0.25)
Controls	Included	Included	Included
Adjusted R^2	.790	.690	.900
Fixed effects	Y, F	Y, F	Y, F
No. of obs.	21,922	21,922	21,922

This table presents the results from the regression of firm investments on the interaction of Tobin's q with nonmarket and market ETF ownership. Market ETFs include those ETFs that physically track broad market indexes, including S&P 500, S&P 1500, Russell 1000, Russell 3000, and the NYSE/NASDAQ Composite Index. Nonmarket ETF ownership is defined as firm ownership by ETFs not classified as Market ETFs. Both firm and year fixed effects are included. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. See Appendix B for variable definitions. *$p<.1$; **$p<.05$; ***$p<.01$.

4. Market versus Nonmarket ETFs

Recent evidence has indicated that ETFs that track market-wide indexes like the S&P 500 are dominated by noise traders, whereas ETFs that are less diversified focusing on certain stocks bring fundamental information into prices (e.g., Bhojraj, Mohanram, and Zhang 2020; Huang, O'Hara, and Zhong 2021). In addition, since market index prices are readily available, we would expect a learning channel to operate via ETFs that do not mimic such indexes. Motivated by these observations, we examine whether it is in fact ownership by nonmarket ETFs that facilitates managerial learning of information from prices.

4.1 Investment-q sensitivity and nonmarket ETF ownership

For the first test, we reexamine our baseline result by including separate variables in our model for ownership by market and nonmarket ETFs, as well as their interaction with Tobin's q. We define market ETFs as those physically tracking broad market indexes, specifically, S&P 500, S&P 1500, Russell 1000, Russell 3000, and the NYSE/Nasdaq Composite Index, and nonmarket ETFs as those do not track such indices. The results, which are shown in Table 4, show that the coefficient between nonmarket ETF ownership and Tobin's q is positive and statistically significant for all three investment measures, whereas the corresponding coefficient for market ETF ownership is insignificant throughout. In terms of economic significance, our estimates imply that a one-standard-deviation increase in Tobin's q (1.30) is associated with an increase of 7.0 (7.7) percentage points in corporate investment (*CAPXRND*) among firms in the bottom (top) quartile of nonmarket ETF ownership. This relative increase in investment represents a change of 6.5% relative to average investments in our sample. The corresponding change for market ETFs is much smaller at 2.8%, which further suggests that information transmitted in stock prices by nonmarket ETFs is more instrumental for guiding investment policy. Overall, the findings in Table 4 indicate that managerial learning from stock prices is facilitated by nonmarket ETFs.[31]

To investigate whether the effect of nonmarket ETFs on investment-q sensitivity is causal, we use inclusion in an industry ETF as a shock to a firm's nonmarket ETF ownership, following the analysis of Huang, O'Hara, and Zhong (2021). Specifically, we match a stock that is included as a member of an industry ETF for the first time to a nonmember stock from the same industry (Fama and French 12-industry classification) using the one-to-one nearest neighbor propensity score matching method. To estimate the propensity score for stocks' industry ETF membership, we estimate a logit model where the dependent variable is a dummy that equals one for member stocks. Matching variables include the log of market capitalization, the log of book-to-market ratio, institutional ownership, analyst coverage, turnover, and idiosyncratic volatility prior to the inclusion event, as in Huang, O'Hara, and Zhong (2021).[32]

Using the matched sample, we then estimate a *diff-in-diff* model around the dates when stocks are added for the first time to industry ETFs. The model estimates the change in the investment-q sensitivity for treatment and control firms in the window 3 years before to 3 years after a stock is included in an industry ETF for the first time. In the model, the dummies $Treat$ and $Post$ equal unity for the treated firms and for the post-inclusion period, respectively. Columns 1–3 in panel A of Table 5 show that the coefficient of interest, $Q_{it-1} * Treat * Post$, is positive and significant, suggesting that the investment-q sensitivity of member stocks increases after their inclusion in industry ETFs, relative to the stocks in the control sample. In Table IA.7, we conduct dynamic analysis, and find that the interaction between Tobin's q, the treatment dummy, and time dummies that flag the years before the firm was included in the nonmarket ETF are statistically insignificant. This alleviates concerns that our result is driven by pervasive differences in investment-price sensitivities between industry ETF member stocks and nonmember stocks. In columns 4–6 in Table 5, we conduct the *diff-in-diff* test using as the relevant event the inclusion of a firm in a market ETF, and find that in this case

31 In Table IA.6, we present various robustness checks for the result in Table 4 (analogous to Table IA.1) and find that in all cases the coefficient between nonmarket ETF ownership and Tobin's q is positive and statistically significant, whereas the corresponding coefficient for market ETF ownership is generally insignificant. We also include a robustness check that stratifies the sample by industry risk exposure, per Huang, O'Hara, and Zhong (2021), and find that our results prevail for both subsamples.

32 Following Huang, O'Hara, and Zhong (2021), our sample of stocks that are added to industry ETFs is focused on those with a market capitalization below the median within the industry. This is because large stocks in an industry ETF cannot be matched with similarly large nonmember stocks from the same industry. We thank Shiyang Huang for sharing the list of industry ETFs with us.

the coefficients for $Q_{i,t-1} * Treat * Post$ are statistically insignificant across all investment measures.

In panel B of Table 5, we estimate the *diff-in-diff* model related to industry ETFs, while decomposing q into its systematic ($Sys\_Q_{i,t-1}$) and idiosyncratic components ($Firm\_Q_{i,t-1}$).[33] As shown in columns 1–3 of panel B, Table 5, the coefficient for $Sys\_Q_{i,t-1} * Treat * Post$ is positive and significant across all three investment measures, whereas the coefficients for $Firm\_Q_{i,t-1} * Treat * Post$ are insignificant throughout.[34]

4.2 Information flows and inclusion in industry ETFs

For our next test, we investigate whether ownership by industry ETFs transmits common information into stock prices, by examining market reactions to earnings announcements. If ownership by industry ETFs stimulates incorporation of common information, the reaction to the earnings surprises, and in particular their common component, should be smaller for stocks after they are added to industry ETFs.[35] Further, ownership by market ETFs should not materially affect the reaction of stock prices to earnings surprises.

To test these ideas, we first calculate standardized unexpected earnings (*SUE*) as the change in split-adjusted quarterly earnings per share from its value four quarters ago, divided by the standard deviation of this change over the prior eight quarters (with a minimum requirement of six quarters of data). We then decompose *SUE* into systematic and firm-specific components ($Systematic\ SUE$ and *Firm SUE*, respectively), using a procedure developed by Jackson, Plumlee, and Rountree (2018) (described in Appendix B). We then use the *diff-in-diff* setting from Table 5 and run the following regression:

$$\begin{aligned} CAR(0,1)_{i,t} &\\ &= \alpha + d_1 * Systematic\ SUE_{i,t} * Treat * Post + d_2 * Firm\ SUE_{i,t} * Treat * Post \\ &\quad + \kappa * Controls + \chi_i + \alpha_t + \epsilon_{i,t}. \end{aligned} \tag{11}$$

The dependent variable above is the 2-day cumulative abnormal return ($CAR(0,1)$), where day 0 is the earnings announcement date. The control variables include the natural logarithm of firm market value at the end of quarter $t-1$, and lagged book-to-market ratios, as well as residual institutional ownership (orthogonalized with respect to ETF ownership). We also control

[33] The common component of Tobin's q for each firm is the fitted value from regressions of firm-level Tobin's q on the aggregate market and industry q (defined at the two-digit SIC level). The firm-specific component is the residual from the above regression.

[34] We perform F-tests for whether the coefficients for the variables that include q are jointly significant in both panels of Table 5. We find joint significance for *CAPXRND* and *RND* within columns 1–3 of each panel.

[35] Several papers present an interpretation of the magnitude of the price reaction around public earnings announcements as an inverse proxy for price efficiency about earnings information; recent examples are Lee and Watts (2021) and Kahraman (2021).

Table 5
Industry ETF inclusion effect on investment-*q* sensitivity

A. Diff-in-diff regression in the matched sample

	Nonmarket ETFs			Market ETFs		
	$CAPXRND_{it}$ (1)	$CAPX_{it}$ (2)	RND_{it} (3)	$CAPXRND_{it}$ (4)	$CAPX_{it}$ (5)	RND_{it} (6)
$Q_{it-1} \times Treat \times Post$	0.008**	0.003**	0.005***	−0.001	−0.002	−0.000
	(2.66)	(2.33)	(2.65)	(−0.30)	(−1.03)	(−0.15)
$Treat \times Post$	−0.011**	−0.006	−0.006***	−0.001	0.002	−0.002
	(−2.27)	(−1.62)	(−2.59)	(−0.12)	(0.49)	(−0.39)
$Q_{it-1} \times Post$	−0.003**	−0.001*	−0.002*	0.001	0.000	0.002
	(−2.14)	(−1.71)	(−1.78)	(0.44)	(0.22)	(0.90)
$Q_{it-1} \times Treat$	−0.004	−0.002	−0.003	−0.008*	−0.003	−0.004
	(−0.84)	(−1.06)	(−0.75)	(−1.88)	(−1.01)	(−1.26)
Q_{it-1}	0.051**	−0.009	0.058***	0.026*	−0.004	0.027**
	(2.40)	(−0.66)	(3.02)	(1.73)	(−0.59)	(2.19)
Controls	Included	Included	Included	Included	Included	Included
Adjusted R^2	.820	.650	.913	.839	.717	.930
Fixed effect	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	2,326	2,326	2,326	6,161	6,161	6,161

B. Decompose Tobin's* q *into systematic and firm-specific components

	$CAPXRND_{it}$ (1)	$CAPX_{it}$ (2)	RND_{it} (3)
$Sys\_Q_{it-1} \times Treat \times Post$	0.010***	0.003*	0.007***
	(3.79)	(1.85)	(3.86)
$Firm\_Q_{it-1} \times Treat \times Post$	0.006	0.001	0.005
	(0.91)	(0.34)	(0.93)
$Treat \times Post$	−0.014**	−0.008*	−0.007**
	(−2.32)	(−1.80)	(−2.27)
$Sys\_Q_{it-1} \times Post$	−0.006***	−0.001	−0.005**
	(−2.72)	(−1.16)	(−2.29)
$Firm\_Q_{it-1} \times Post$	−0.001	−0.001	0.000
	(−0.19)	(−0.31)	(0.20)
$Sys\_Q_{it-1} \times Treat$	−0.001	−0.013	0.011
	(−0.07)	(−1.62)	(1.41)
$Firm\_Q_{it-1} \times Treat$	−0.003	−0.000	−0.002
	(−0.43)	(−0.35)	(−0.48)
Sys_Q_{it-1}	0.015	0.005	0.009**
	(1.41)	(0.94)	(2.26)
$Firm\_Q_{it-1}$	0.019***	0.002	0.016***
	(3.86)	(0.45)	(3.04)
Controls	Included	Included	Included
Adjusted R^2	.819	.653	.913
Fixed effects	Y, F	Y, F	Y, F
No. of obs.	2,322	2,322	2,322

This table reports results from estimating the effect of stocks' inclusion into industry ETFs on investment-*q* sensitivity. Following Huang et al. (2021), when a stock is included in an industry ETF for the first time, we match this member stock with a nonmember stock from the same industry (Fama and French 12-industry classification) using the one-to-one nearest-neighbor propensity score matching method. To estimate the propensity score for the industry ETF constituent, we estimate a logit model where the dependent variable is a dummy that equals one for the member stock and zero for the nonmember stock. Matching variables include the logarithm of market capitalization (*log(ME)*), the logarithm of the book-to-market ratio (*log(BM)*), institutional ownership (*IO*), analyst coverage *(# analysts)*, turnover (*Turnover*), and idiosyncratic volatility (*IVOL*) prior to the inclusion event. We focus on member stocks with a market capitalization below the median within the industry since a large stock in the industry ETF cannot be matched with a similarly large nonmember stock from the same industry. Columns 1 to 3 of panel A report the results from estimating difference-in-differences models of investment-*q* sensitivity for 3-year windows around the inclusion of stocks in industry ETFs in the matched sample. *Post* is a dummy variable that equals one for the period after inclusion in industry ETFs, and zero otherwise. *Treat* is a dummy that equals one for a firm included for the first time in an industry ETF, and zero for the matched control firms. *Post* and *Treat* are not included in the model as they are subsumed by firm and year fixed effects, respectively. In columns 4 to 6 of panel A, we report the difference-in-differences estimation results, using stocks' first-time inclusion in market ETFs as placebo events. In panel B, we further decompose each firm's Tobin's q into systematic and firm-specific components and reestimate the difference-in-differences regression. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. $*p<.1$; $**p<.05$; $***p<.01$.

Table 6
Industry ETF inclusion effect on market reactions to earnings announcements

	Industry ETFs *[CAR(0,1)]* (1)	Market ETFs *[CAR(0,1)]* (2)
Systematic SUE_{it} × *Treat* × *Post*	−0.015***	−0.002
	(−2.74)	(−0.39)
Firm SUE_{it} × *Treat* × *Post*	−0.001	−0.004
	(−0.34)	(−1.52)
Systematic SUE_{it} × *Treat*	0.004	0.001
	(0.84)	(0.31)
Firm SUE_{it} × *Treat*	−0.000	0.002
	(−0.08)	(1.05)
Systematic SUE_{it} × *Post*	0.007*	0.001
	(1.96)	(0.29)
Firm SUE_{it} × *Post*	0.001	0.005*
	(0.38)	(1.91)
Treat × *Post*	0.00	0.001
	(0.08)	(0.42)
Systematic SUE_{it}	0.012***	0.017***
	(3.51)	(5.57)
Firm SUE_{it}	0.014***	0.015***
	(6.62)	(8.68)
Controls	Included	Included
Adjusted R^2	.051	.105
Fixed effects	Q, F	Q, F
No. of obs.	9,057	24,648

This table reports results from estimating the effect of stocks' inclusion in industry ETFs on market reactions to earnings announcements. The dependent variable in the model is the cumulative abnormal return [*CAR(0,1)]* over a 2-day window around quarterly earnings announcement, where day 0 is the earnings announcement date. The abnormal return is calculated as the raw stock return minus the value weighted CRSP index return. *Standardized unexpected earnings* (*SUE*) is defined as the change in split-adjusted quarterly earnings per share from its value four quarters ago, divided by the standard deviation of this change in quarterly earnings over the prior eight quarters (with a minimum requirement of six quarters). We decompose the concurrent *SUE* for each firm-quarter into *Systematic SUE* and *Firm SUE*, as described in Appendix B. *Post* is a dummy variable that equals one for the period after first-time inclusion in industry ETFs, and zero otherwise. *Treat* is a dummy that equals one for the firm that is included for the first time in an industry ETF, and zero for the matched control firms, following the procedure explained in Table 5. Column 2 reports the difference-in-differences estimation results, using stocks' inclusion in market ETFs as placebo events. We include time fixed effects (fiscal year-quarter) and firm fixed effects in the regression. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. *$p<.1$; **$p<.05$; ***$p<.01$.

for the interaction of residual institutional ownership with the two components of earnings surprises.

Column 1 of Table 6 shows the results. We find that the coefficient for $Systematic\ SUE_{i,t} * Treat * Post$ is negative and statistically significant, while the coefficient for the interaction between $Firm\ SUE_{i,t} * Treat * Post$ is not. This result is in line with our previous findings in Section 4.1, showing that ownership by industry ETFs brings systematic information into stock prices. The estimates in Table 6, column 1, imply that a one-standard-deviation increase in the common component of *SUE* is associated with a decrease of 1.40% in $CAR(0,1)$ around earnings announcements after a stock is included in an industry ETF for the first time relative to stocks not included in such ETFs.[36] In

[36] Again, we verify via an F-test that the coefficients in front of the two *SUE* components are jointly significant.

Table IA.8, we conduct dynamic analysis, and find that the interaction between the systematic *SUE* with the treatment dummy and time dummies for the period before the firm's inclusion in the industry ETF, are statistically insignificant. This alleviates concerns that our results are driven by pervasive differences in stock price reactions to earnings surprises across stocks that do and do not belong to industry ETFs.

In column 2 of Table 6, we conduct the *diff-in-diff* test when the relevant event is the inclusion of a firm in a market ETF. In this case we find that the relevant interaction coefficients are statistically insignificant, in line with the notion that it is nonmarket ETFs that bring common information into stock prices.[37]

4.3 ETF ownership and investment sensitivity to peers' prices

Proposition 3 of our model predicts that since ETFs help transmit common information, managers rely more on own prices and less on peers' prices in their investment decisions when ownership by nonmarket ETFs is high. Therefore, the real investment of firms with higher (lower) nonmarket ETF ownership should be less (more) responsive to peers' q. To test this prediction, we augment the baseline regression by including the average q of peer firms (PQ_{it-1}) and its interaction with the firm's ownership by nonmarket ETFs ($PQ_{it-1} * NonMktETF_{it-1}$) and market ETFs ($PQ_{it-1} * MktETF_{it-1}$). Our model predicts a negative coefficient for $PQ_{it-1} * NonMktETF_{it-1}$, which captures the impact of nonmarket ETF activity on the investment sensitivity to peers' prices. Following the literature, we use the Text-Based Network Industry Classification (TNIC) approach to identify peer firms, as developed by Hoberg and Phillips (2010, 2016).[38]

Panel A of Table 7 reports the results. In the first six columns, we present the baseline coefficients of own and peers' q when these are, in turn, the only variables in the regressions where the dependent variables are the three measures of investment. Consistent with Dessaint et al. (2019), the coefficients of both the variables are positive and generally significant. When we add controls, and ETF ownership and its interactions with q, we find that in line with our previous findings, the coefficient for $Q_{it-1} * NonMktETF_{it-1}$ is positive and significant across all three investment measures, whereas the coefficient for $Q_{it-1} * MktETF_{it-1}$ is insignificant throughout. Consistent with Proposition 3,

[37] This is the direct effect of ETF ownership discussed in Section 1. In our model, an indirect effect of such ownership increases the flow of firm-specific information. In this regard, note that in both Table 5, panel B, and Table 6, column 1, the sign of the coefficient between the interaction of the idiosyncratic component of the variable of interest (Tobin's q or *SUE*) and $Treat * Post$ indicates an improvement in price informativeness after a firm is included in a nonmarket ETF. Therefore, even though this effect is not statistically significant, our findings do not contradict those in Huang, O'Hara, and Zhong (2021), who propose a different mechanism: that industry ETFs help facilitate the flow of firm-specific information via better hedging opportunities for informed investors.

[38] The data can be obtained from http://hobergphillips.usc.edu/idata/Readme_tnic3.txt. We use TNIC to identify peers, as this measure captures firms' product market spaces in a more timely manner, whereas generic industry identifiers are significantly more outdated.

Table 7
ETF ownership and investment sensitivity to peers' stock prices

***A. Investment sensitivity to peers'* Q**

	$CAPXRND_{it}$ (1)	$CAPX_{it}$ (2)	RND_{it} (3)	$CAPXRND_{it}$ (4)	$CAPX_{it}$ (5)	RND_{it} (6)	$CAPXRND_{it}$ (7)	$CAPX_{it}$ (8)	RND_{it} (9)
Q_{it-1}	0.019***	0.009***	0.009***				0.057***	0.011***	0.043***
	(15.74)	(15.08)	(8.95)				(8.03)	(3.54)	(7.82)
PQ_{it-1}				0.004***	0.005***	−0.001	0.003	0.003**	0.001
				(2.91)	(4.91)	(−0.98)	(1.37)	(2.28)	(0.32)
$Q_{it-1} \times NonMktETF_{it-1}$							0.169***	0.061***	0.125***
							(3.57)	(2.81)	(3.45)
$PQ_{it-1} \times NonMktET_{Fit-1}$							−0.110**	−0.023	−0.089**
							(−2.17)	(−0.94)	(−2.47)
$Q_{it-1} \times MktETF_{it-1}$							0.160	−0.020	0.133
							(1.22)	(−0.33)	(1.33)
$PQ_{it-1} \times MktETF_{it-1}$							0.079	−0.036	0.064
							(0.47)	(−0.41)	(0.59)
$NonMktETF_{it-1}$							−0.294***	−0.138**	−0.172**
							(−2.64)	(−2.21)	(−2.33)
$MktETF_{it-1}$							0.110	0.332*	−0.061
							(0.31)	(1.69)	(−0.27)
Controls							Included	Included	Included
Adjusted R^2	.771	.679	.876	.757	.668	.872	.789	.692	.900
Fixed effect	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	20,084	20,084	20,084	20,084	20,084	20,084	20,084	20,084	20,084

(*Continued*)

Table 7
Continued

B. Decomposing firm's own* Q *and peers'* Q *into noise and fundamental components

	First-stage		Second-stage		
	Q_{it-1} (1)	PQ_{it-1} (2)	$CAPXRND_{it}$ (3)	$CAPX_{it}$ (4)	RND_{it} (5)
$MFflow_{it-1}$	3.519***				
	(4.33)				
$PMFflow_{it-1}$		31.885***			
		(14.90)			
$Q\_fundamental_{it-1} \times NonMktETF_{it-1}$			0.162**	0.055***	0.124***
			(2.57)	(2.92)	(3.34)
$Q\_Noise_{it-1} \times NonMktETF_{it-1}$			0.749***	0.280**	0.511***
			(3.52)	(2.04)	(4.30)
$Q\_fundamental_{it-1} \times MktETF_{it-1}$			0.200	0.012	0.158
			(1.60)	(0.13)	(1.56)
$Q\_Noise_{it-1} \times MktETF_{it-1}$			−0.209	0.436	−0.576
			(−0.20)	(0.68)	(−1.13)
$PQ\_fundamental_{it-1} \times NonMktETF_{it-1}$			−0.136**	−0.037	−0.101***
			(−2.42)	(−1.22)	(−3.64)
$PQ\_Noise_{it-1} \times NonMktETF_{it-1}$			−0.306**	−0.124	−0.181**
			(−2.03)	(−1.23)	(−1.97)
$PQ\_fundamental_{it-1} \times MktETF_{it-1}$			0.167	0.020	0.090
			(1.27)	(0.13)	(0.86)
$PQ\_Noise_{it-1} \times MktETF_{it-1}$			−0.859	−1.263*	0.258
			(−0.92)	(−1.84)	(0.63)
$Q\_fundamental_{it-1}$			0.058***	0.010***	0.046***
			(6.12)	(2.92)	(8.54)
Q_Noise_{it-1}			0.054***	0.004	0.045***
			(2.60)	(0.31)	(3.34)
$PQ\_fundamental_{it-1}$			0.003	0.003	0.000
			(0.85)	(1.12)	(0.23)
PQ_Noise_{it-1}			0.018	0.019	0.002
			(1.03)	(1.67)	(0.25)
$NonMktETF_{it-1}$			−0.920***	−0.309**	−0.678***
			(−3.97)	(−2.18)	(−3.08)
$MktETF_{it-1}$			2.819***	2.084***	0.882***
			(3.35)	(2.82)	(3.07)
Controls	No	No	Included	Included	Included
F-test	18.75	222.01			
Adjusted R^2	.662	.709	.788	.692	.899
Fixed effect	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	19,731	19,731	19,731	19,731	19,731

Panel A reports the results from the regression that includes the average Tobin's q of firm i's peers (PQ_{it-1}), and its interaction with firm i's ownership by market ETFs ($PQ_{it-1} \times MktETF_{it-1}$) and nonmarket ETFs ($PQ_{it-1} \times NonMktETF_{it-1}$). In panel B, we decompose firms' own Tobin's q and peers' q into noise and fundamental components and interact with nonmarket and market ETF ownership separately. Following Dessaint et al. (2018), we regress Tobin's q on stock-level fund outflow induced price pressure measure (*MFflow*) and obtain the regression fitted value (residual) as the *Q_noise* (*Q_fundamental*). *PQ_noise* and *PQ_fundamental* are constructed in a similar way. We follow the Dessaint et al. (2021) approach to measure *MFflow*. Both firm and year fixed effects are included. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. See Appendix B for variable definitions. $*p<.1$; $**p<.05$; $***p<.01$.

the coefficient for the interaction $PQ_{it-1} * NonMktETF_{it-1}$ is negative across all three investment measures, and statistically significant for two of them. In contrast, the coefficient for $PQ_{it-1} * MktETF_{it-1}$ is insignificant throughout.

Although in our model ETFs can increase the incorporation of both firm-specific and common information, we include a test for the differential impact of the types of information. In Table IA.9 within the Internet Appendix, we decompose the q's of a firm and its peers into their systematic and idiosyncratic components. We find that the interaction of the systematic component of peers' q with nonmarket ETF ownership is stronger than for the firm-specific component. Specifically, whereas the coefficients for $Firm\_PQ_{it-1} * NonMktETF_{it-1}$ are statistically insignificant, the coefficient for $Sys\_PQ_{it-1} * NonMktETF_{it-1}$ is negative and significant for two of three investment measures.

Next, we recognize the desirability of including a test of whether prices provide useful information to managers, as opposed to reflecting what managers already know. For this purpose, an appropriate test is to split q into a component that is unlikely to be a part of the managers' private information set, and an orthogonal component. To this end, we conduct a test based on Dessaint et al. (2019) (see also Zuo 2016). We follow the procedure in Edmans, Goldstein, and Jiang (2012), and calculate the variable *MFflow* for each firm-quarter, which measures the price pressure related to large outflows of capital from mutual funds that hold the stock. The specific *MFflow* measure we use is calculated using the technique of Dessaint et al. (2021).[39] Edmans, Goldstein, and Jiang (2012) and Dessaint et al. (2019) propose that because large negative values of *MFflow* are contemporaneously related to negative stock returns, which reverse in the near future, *MFflow* reflects noise trading, as opposed to trading on fundamental information. Because it is unlikely that managers know in advance what noise traders who invest in mutual funds will do, this test can be used to examine whether managers condition on prices when choosing their investment levels. As shown in Figure IA.3 in the Internet Appendix, *MFflow* is contemporaneously related to large negative returns, which completely revert in the following months.[40] Thus, this key pattern shown by Edmans, Goldstein, and Jiang (2012) and Dessaint et al. (2019) also emerges in our sample, which validates the measure.

We decompose both peers' and own q into noise- and fundamental-related components by annually regressing q on *MFflow*, and calculating the predicted component and the residual. As shown in the first two columns of Table 7, panel B, in the first-stage regression, where the dependent variable is, in turn, own and peers' q , the coefficients for *MFflow* are positive and significant, consistent with the results reported by Dessaint et al. (2019). We present results analogous to panel A using the components of q within columns 3–5 of Table 7, panel B. We find evidence of managerial learning consistent with Proposition 3, in that the reliance on peers' q declines as nonmarket ETF ownership rises,

[39] This technique addresses an issue in Wardlaw (2020). Specifically, the concern is that the denominator (scale factor) for the original measure (total dollar volume) involves a market price, and Dessaint et al. (2021) use the price as of the end of the previous quarter to avoid a mechanical relation between fund flows and current returns.

[40] To construct Figure IA.3, we define an "event" as a firm-quarter in which *MFflow* falls below the 10th percentile value of the full sample. We then trace the cumulative abnormal returns (CAR) over the CRSP equal-weighted or value-weighted index from 15 months before the event to 24 months after.

for both components of q. There also is evidence that managerial reliance on one's own q increases for both components of q, with rising nonmarket ETF ownership. There is no corresponding evidence for market ETF ownership. As Dessaint et al. (2019) point out, while the fundamental component of q may include overlap between managerial and price-related information, the *MFflow* channel is unique to learning from prices, given that managers cannot fully disentangle noise from information when conditioning on prices.[41] Hence, overall, the evidence in Table 7 supports the learning channel.[42]

4.4 Individual stock prices versus ETF prices as conditioning variables

A question that relates to our hypothesis is whether managers primarily rely on their own stock prices or ETF prices to extract common information. First, in our model, we propose that ETFs, by facilitating the incorporation of common information about ζ, stimulate information collection about product demand or cost structure (β), where β is solely revealed through stock prices. This incentivizes managers to condition on the stock price as a single, easily available number that aggregates information on both ζ and β, as well as the idiosyncratic components of value (captured by θ). Second, as long as ETF prices are noisy, they are not perfect substitutes for stock prices.[43] Third, we propose that managers condition largely or exclusively on stock prices because they are much more salient to managers than scores of unfamiliar ETF prices.[44] Thus, for managers, conditioning on several ETFs is more cognitively challenging than conditioning on own stock prices. This argument is in the spirit of other papers. For example, Hirshleifer and Teoh (2003, p. 339) state that "information that is presented in salient, easily processed form is assumed to be absorbed more easily than information that is less salient," and Hong, Torous, and Valkanov (2007, p. 371) argue that "investors, due to limited cognitive capabilities, have a hard time processing information from asset markets that they do not participate in." Thus, managers, like investors, may condition on more salient information, that is, on own stock prices, in preference to ETF prices.

To test the above conjecture, we examine whether the effect of nonmarket ETF ownership on investment-q sensitivity is stronger for a firm whose return has a lower average correlation with the return of nonmarket ETFs holding the stock. The rationale is that with low correlation, learning from multiple ETF

[41] Note that the negative coefficient of $PQ_{Noise} \times NonMktETF$ is unique to the mechanism that ETFs facilitate the flow of common nonmarket information to managers. It does not obtain under the conjecture that ETFs faciliate firm-specific information acquisition.

[42] In Table IA.10, we perform the difference-in-differences regression of Table 5 using the Q decomposition of Table 7, and find again that the flow-related component of Q remains significant and of the right sign, for two of the three investment measures.

[43] Provided ETFs incorporate information about factors not relevant to the stock, or additional noise trading, their prices will only be a noisy signal for managers and thus will not supplant own firm prices.

[44] On average a stock in our sample is held by more than 20 ETFs and the maximum number of ETFs holding a stock exceeds 100.

Table 8
Tests of the managerial learning constraints channel

	Low correlatio			High correlation		
	(1)	(2)	(3)	(4)	(5)	(6)
	$CAPXRND_{it}$	***$CAPX_{it}$***	***RND_{it}***	***$CAPXRND_{it}$***	***$CAPX_{it}$***	***RND_{it}***
$Q_{it-1} \times NonMktETF_{it-1}$	0.171***	0.068***	0.113***	0.029	0.008	0.030
	(3.15)	(2.68)	(2.67)	(0.45)	(0.29)	(0.57)
$Q_{it-1} \times MktETF_{it-1}$	0.190	0.003	0.133	0.268	0.017	0.247
	(1.27)	(0.04)	(1.15)	(1.34)	(0.18)	(1.55)
Q_{it-1}	0.051***	0.011***	0.038***	0.046***	0.008	0.036***
	(6.87)	(3.57)	(6.36)	(3.79)	(1.40)	(3.93)
$NonMktETF_{it-1}$	−0.554***	−0.223***	−0.356***	−0.120	−0.014	−0.116*
	(−5.42)	(−3.58)	(−4.80)	(−1.15)	(−0.21)	(−1.76)
MktETF it-1	0.174	0.161	0.071	0.056	0.238	−0.182
	(0.59)	(0.96)	(0.34)	(0.14)	(0.97)	(−0.70)
Controls	Included	Included	Included	Included	Included	Included
Adjusted R^2	.794	.637	.897	.801	.740	.925
Fixed effect	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	10,260	10,260	10,260	10,512	10,512	10,512

This table reports the results from the regressions of firm investments ($CAPXRND_{it}$, $CAPX_{it}$ and RND_{it}) on the interaction of Tobin's q and nonmarket ETF ownership ($Q_{it-1} \times NonMktETF_{it-1}$) and market ETF ownership ($Q_{it-1} \times MktETF_{it-1}$) for subsamples formed based on the average return correlation between the stock returns of firm i and each nonmarket ETF holding stock i (using daily returns from the previous 9 months). Columns 1 to 3 (columns 4 to 6) report the results where the average correlation is below (above) the sample median in year $t-1$. Both firm and year fixed effects are included. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. See Appendix B for variable definitions. *$p<.1$; **$p<.05$; ***$p<.01$.

prices is more challenging and the stock price serves as a better conditioning variable. To conduct this test, we calculate the average correlation between a stock's return and that on the corresponding nonmarket ETFs from the past 9 months of daily returns. We then split the sample into two equal groups based on the average correlation. From Table 8, we find that the coefficient for $Q_{it-1} * NonMktETF_{it-1}$ is indeed larger and more significant when the average correlation is below the median (columns 1 to 3). In contrast, columns 4 to 6 show that the coefficient is not significant for the complementary sample. This result supports the salience hypothesis.[45]

4.5 Cross-sectional heterogeneity

In this section, we consider conditional variation in the effect of nonmarket ETF ownership on the investment-q relation. We focus on the comprehensive measure of investment, *CAPXRND*. The results involving $CAPX$ appear in Table IA.12, and we discuss these at the end of this section.

First, we test the two cross-sectional predictions of our model. Cross-Sectional Prediction 1 in Section 1.6 is that the positive effect of nonmarket

[45] We conduct an additional test by controlling for the manager's learning from ETF prices. Specifically, we add annual ETF-level returns in the baseline regression. We also add the interaction of ETF-level returns (i.e., the equally weighted average returns of all ETFs holding the firm's stock) with the stock's market beta because the effect of ETF prices on investment could be different for stocks with differential exposure to common factors. Table IA.11 of the Internet Appendix shows that the coefficient of the interaction term $Q_{it-1} * NonMktETF_{it-1}$ remains positive and significant.

ETF ownership on investment-q sensitivity is stronger if the common signal's precision is low, because in this case, more investors collect factor information. Note that this precision could be low for two reasons: First, the factor may be more volatile; second, the sensitivity of a firm's cash flow to the factor may be larger. Cross-Sectional Prediction 2 of our model is that the positive effect of nonmarket ETF ownership on investment-q sensitivity is stronger when the manager has more precise firm-specific information. The intuition is that when the manager has extremely noisy firm-specific information, virtually all the Bayesian weight is attached to the stock price. Thus, the sensitivity of investment to the stock price, which is related to the weight, is at its maximum and does not move with ETF ownership. However, when the manager has complementing firm-specific information that is very precise, then the investment-to-price sensitivity is very responsive to the additional factor trading induced by the ETF.

To test the first prediction, we use two measures to capture the precision of common information. First, we use an industry cash flow beta as a proxy for sensitivity of a firm's cash flow to common factors.[46] Second, we use the volatility of industry-level profitability as a proxy for the uncertainty of the factor.[47] For testing the second prediction, we use the profitability of insider trades as a proxy for the precision of managerial firm-specific information, as we show in the model that expected insider trading profits increase with the precision of the manager's private firm-specific information.[48] We include both ownership by market and nonmarket ETFs in the models, but expect that our cross-sectional predictions hold for the latter and not the former.

We define a dummy variable (Dum_{it-1}) that equals one for firm i if its cash flow beta is above the industry median beta (the benchmark is the industry median, because betas differ considerably across industries), and zero otherwise. We define equivalent dummies if the volatility of industry-level profitability, or profitability of insider trades, are above their full-sample medians in year $t-1$, and zero otherwise. We interact this dummy indicator with the product of q and the ETF ownership variables, and investigate the coefficient of these triple interaction terms.

---

[46] Specifically, the cash flow beta is obtained by regressing an individual firm's ROE on value-weighted industry-level ROE (defined at the two-digit SIC code level) and aggregate (market-level) ROE, using the past 5 years of quarterly data (with a minimum of eight observations in the regression). We use the coefficient of industry-level ROE as the firm's cash flow beta.

[47] This volatility is defined as the standard deviation of the annual industry-level profit margin over a 10-year rolling window. The industry-level profit margin is measured as the operating income after depreciation divided by the total sales of all firms within the same industry.

[48] Following the literature, we measure insider trading profitability by the average one-month market-adjusted returns following insiders' net transactions in that month. We obtain insiders' trades from the Thomson Financial Insider Filing database, and, as in other studies (e.g., Dessaint et al. 2019), we restrict our attention to open market stock transactions initiated by the top-five highest-ranked executives (CEO, CFO, COO, President, and Chairman of the board).

Table 9
Cross-sectional heterogeneity tests

	Importance of common information	Uncertainty of common information	Managerial firm-specific information	Uncertainty of information environment		Growth potential
	(1)	(2)	(3)	(4)	(5)	(6)
	Cash flow beta	*PROFVOL*	*Insider profit*	*Analyst coverage*	*Forecast dispersion*	*B/M*
$Q_{it-1} \times NonMktETF_{it-1} \times Dum_{it-1}$	0.089*	0.153**	0.150**	0.166**	0.201***	0.224**
	(1.85)	(1.97)	(2.53)	(2.31)	(2.83)	(2.13)
$Q_{it-1} \times MktETF_{it-1} \times Dum_{it-1}$	0.082	0.025	−0.422**	−0.223	−0.101	−0.670*
	(0.47)	(0.11)	(−2.09)	(−0.94)	(−0.46)	(−1.95)
$Q_{it-1} \times Dum_{it-1}$	−0.006***	−0.005**	0.000	−0.010***	−0.003	−0.013
	(−2.75)	(−2.03)	(0.15)	(−4.37)	(−1.26)	(−0.98)
Q_{it-1}	0.054***	0.057***	0.056***	0.074***	0.066***	0.054***
	(7.62)	(8.29)	(7.84)	(9.15)	(7.89)	(4.17)
Controls	Included	Included	Included	Included	Included	Included
Adjusted R^2	.789	.790	.794	.803	.804	.794
Fixed effect	Y, F	Y, F	Y, F	Y, F	Y, F	Y, F
No. of obs.	20,879	21,919	19,327	18,855	16,786	21,922

This table reports the results from the baseline regression of firm investment ($CAPXRND_{it}$) on the interaction of Tobin's q and nonmarket ETF ownership (Q_{it-1}* $NonMktETF_{it-1}$) and market ETF ownership ($Q_{it-1} \times MktETF_{it-1}$) conditional on the importance of common information (column 1), the uncertainty of common information (column 2), and the precision of managerial firm-specific information (column 3), respectively. The importance of common information is measured using a stock's industry cash flow beta. The uncertainty of common information is measured as the volatility of industry-level profitability (*PROFVOL*). Managerial firm-specific information is measured by the average profitability of insider trading (*InsiderProfit*) for each firm over the past 3 years. For the first partitioning variable, we create a dummy equal to one if its value is above the median of the industry in year $t-1$. For the other two partitioning variables in columns 2 and 3, we create a dummy that equals to one if their value is above the median in the whole sample in year $t-1$. In columns 4 and 5, we define the partitioning dummy based on firms' information environments, and in column 6 based on book-to-market ratio. Information environment is measured using analyst coverage and analyst earnings forecast dispersion. We define $Dum = 1$ if firm i is above the sample median in terms of analyst forecast dispersion, and below median in terms of analyst coverage in year $t-1$. For book-to-market (B/M) ratio, we define $Dum = 1$ if a firm's B/M ratio is below the industry median in year $t-1$. Both firm and year fixed effects are included. t-statistics, reported in parentheses, are based on standard errors clustered at the firm level. See Appendix B for variable definitions. $*p<.1$; $**p<.05$; $***p<.01$.

The results, shown in Table 9, columns 1–3, are consistent with our predictions. In columns 1 and 2, we find that the coefficient between $Q_{it-1} * NonMktETF_{it-1} * Dum_{it-1}$ is positive and significant, showing that the effect of ownership by nonmarket ETFs on investment-q sensitivity is larger among firms with higher cash flow betas, or firms operating in industries where shocks to profitability are more uncertain, respectively. Column 3 shows that the effect of nonmarket ETFs on investment-q sensitivity is more pronounced when managers are likely to possess more precise firm-specific information.

In columns 4–6 of Table 9, we test additional cross-sectional predictions that, although not directly predicted by our model, are intuitively suggested by the managerial learning hypothesis. The channel predicts that the impact of nonmarket ETFs should be greater for firms with more challenging information environments, as ETFs should improve such stocks' informational efficiency the most. We measure information environments using analysts' coverage and forecast dispersion, and find support for this hypothesis, as the effect of nonmarket ETFs on investment-q sensitivity is larger among firms with lower coverage and higher forecast dispersion, respectively.[49]

Finally, in column 6 of Table 9, we conduct a cross-sectional test on the interaction between growth opportunities and investment-q sensitivity. The rationale is that managers should rely more on stock prices for information when growth potential of their firm is high. Thus, we define Dum_{it-1} based on the firms' market-book ratio as a proxy for growth opportunities, and estimate our baseline model. We find that the coefficient for the triple interaction $Q_{it-1} * NonMktETF_{it-1} * Dum_{it-1}$ is positive and significant, showing that the effect of nonmarket ETF ownership on investment-q sensitivity is larger among firms with higher growth potential.

As mentioned above, Table 9 considers *CAPXRND*, which includes both traditional investment and *RND*. We present results for the measure of real investment that excludes R&D expenditures (i.e., $CAPX$) in Table IA.12. The table indicates that while the coefficient signs point in the same direction as those in Table 9, the significance of the cross-sectional heterogeneity discussed above obtains only for the cuts corresponding to industry-level profitability volatility and growth/value. We propose that the importance of R&D in managerial learning varies across the cuts, thus implying varying levels of significance for the heterogeneity. Overall, the results for the comprehensive measure *CAPXRND* suggest that the positive effect of nonmarket ETF ownership on investment-q sensitivity is stronger in cases where managers are more likely to rely on stock prices for information.

49 We measure forecast dispersion (analyst coverage) as the monthly average of the coefficient of variation of annual EPS forecasts (number of analysts who issue forecasts) in year $t-1$, using data from the IBES summary files (see Appendix B for specific definitions).

5. Alternative Explanations

In this section we go beyond our theoretical setting to conduct tests that address alternative explanations for our findings. For brevity, we insert the relevant results in the Internet Appendix.

5.1 Improvement in corporate governance?

Appel, Gormley, and Keim (2016) show that increases in ownership by passive investors improve the quality of firms' governance.[50] Improved governance could increase investment-q sensitivity by better aligning managers' interests with those of shareholders (John, Litov, and Yeung 2008). If this mechanism drives our findings, we should find a large effect of ETF ownership on investment-q sensitivity for firms that have weak governance to begin with, since for such firms the ETF-related improvement in governance would have greater impact.

To test this possibility, we partition firms into subsamples (strong and weak governance) based on firm-level governance indexes at year $t-2$ (with the dependent variable measured in year t). The G-index (E-index) is constructed by adding one point for each of the 24 (6) (anti)takeover provisions listed in Gompers, Ishii, and Metrick (2003). Higher values imply weaker governance.[51]

The results in Table IA.13 show that the coefficient for the interaction of q with nonmarket ETF ownership is positive and significant only for the subsamples with strong governance, for both measures of governance in the case of *CAPXRND* and for one measure in the case of $CAPX$. This finding does not accord with the governance improvement channel. Rather, overall, the result suggests that good governance is necessary for managers to use the information contained in prices for making real investment decisions.

5.2 Relaxed financial constraints?

Firms with higher nonmarket ETF ownership might have easier access to external finance, and thus face more relaxed financial constraints. Less binding constraints strengthen investment-q sensitivity by allowing managers to promptly adjust investments in response to price signals (e.g., Bakke and Whited 2010). In this subsection, we examine whether ETFs are indeed associated with lower constraints. Specifically, we regress six different measures of financial constraints of firm i at year t on its ETF ownership at year $t-1$. These measures are the text-based ones developed by Hoberg and Maksimovic (2015) (HM), the change in the credit default spread (ΔCDS, from Markit) as a measure of the cost of debt, the firms' total payout ratio (dividends

50 However, other studies have suggested a reverse effect of passive institutions on governance (e.g., Schmidt and Fahlenbrach 2017; Bebchuk, Cohen, and Hirst 2017).

51 Governance is strong when the value of the G-index (E-index) is below 10 (3); otherwise, it is defined as weak. For this test we use data from 2004 to 2009 because the G and E indexes are not available after 2008.

+ repurchases), and equity and debt issuance. The tests are performed for the subset of firms for which the constraint measures are available. The results reported in Table IA.14 show that higher nonmarket ETF ownership is not significantly associated with the HM measures, is associated with lower payout ratios and reduced equity issuance, but bears no relation to changes in credit default spreads or debt issuance. Thus, overall, the link between ownership by nonmarket ETFs and financial constraints is tenuous.

6. Concluding Remarks

We examine the link between ETF ownership and real investment. On the one hand, ETF ownership might reduce price informativeness by increasing noise trading. On the other hand, ETFs might facilitate the incorporation of industry and sector-related information into stock prices. We present empirical results that validate the latter notion. Specifically, higher nonmarket ETF ownership is associated with an increased sensitivity of corporate investment to stock prices, which is consistent with managerial learning. In additional tests, we find that investment-q sensitivity rises, and common information is more promptly incorporated into prices, for stocks that are added to industry ETFs. Overall, our evidence accords with the view that nonmarket ETFs can help contribute to real efficiency via the production of industry and sector information.

Appendix A

A.1. Proofs

Proof of Lemma 1: The price function is given by

$$p\left(\sum_{n_1} x_j+\sum_{n_2} y_j+\sum_{n_3} z_j+e\right)=\lambda_1\left(\sum_{n_1} x_j+\sum_{n_2} y_j+\sum_{n_3} z_j+e\right)$$

$$=\lambda_1\begin{pmatrix} n_1(\gamma_1\zeta) \\ +n_2(\eta_1\beta) \\ +n_3(\kappa_1\theta) \\ +e \end{pmatrix},$$

where $n_1 \geq 1$, $n_2 \geq 1$, and $n_3 \geq 1$. For a trader informed about ζ, the trading strategy is given by

$$\max_x \mathbb{E}[x(v-p)|\zeta]$$

$$\Rightarrow \max_x \mathbb{E}\left[x\left(v-\left[\lambda_1\left(x+\sum_{n_1-1} x_j+\sum_{n_2} y_j+\sum_{n_3} z_j+e\right)\right]\right)\middle|\zeta\right]$$

which implies

$$x=\frac{\zeta-\lambda_1(n_1-1)\gamma_1\zeta}{2\lambda_1}, \tag{A.1}$$

$$\mathbb{E}(\pi_1|\zeta)=\lambda_1\cdot x^2. \tag{A.2}$$

Similarly, for a trader informed about β, we have

$$y=\frac{\beta-(n_3-1)\kappa_1\theta}{2\lambda_1} \tag{A.3}$$

$$\mathbb{E}(\pi_3|\theta)=\lambda_1\cdot z^2. \tag{A.4}$$

Comparing Eq. (A.1) with $x=\gamma_1\zeta$ yields

$$\gamma_1=\frac{1}{\lambda_1(n_1+1)}$$

The ex ante expected profit for the first type of trader can be expressed by

$$\mathbb{E}(\pi_1)=\mathbb{E}\left(\lambda_1 x^2\right)=\frac{\tau_\zeta^{-1}}{\lambda_1(n_1+1)^2},$$

where $\mathbb{E}(\cdot)$ is the unconditional expectation operator over ζ. Similarly, we can work out

$$\eta_1=\frac{1}{\lambda_1(n_2+1)};\ \kappa_1=\frac{1}{\lambda_1(n_3+1)},$$

and

$$\mathbb{E}(\pi_2)=\frac{\tau_\beta^{-1}}{\lambda_1(n_2+1)^2};\ \mathbb{E}(\pi_3)=\frac{\tau_\theta^{-1}}{\lambda_1(n_3+1)^2},$$

Now we move to solve the market maker's problem. The pricing rule is given by

$$\mathbb{E}\left(v|\sum_{n_1}x_j+\sum_{n_2}y_j+\sum_{n_3}z_j+e\right)=\mathbb{E}\left(\zeta+\beta+\theta|\begin{pmatrix} n_1\gamma_1\zeta \\ +n_2\eta_1\beta \\ +n_3\kappa_1\theta \end{pmatrix}+e\right)$$

which implies

$$\lambda_1=\frac{\operatorname{cov}\left(\zeta+\beta+\theta,\begin{pmatrix} n_1\gamma_1\zeta \\ +n_2\eta_1\beta \\ +n_3\kappa_1\theta \end{pmatrix}+e\right)}{\operatorname{var}\left(\begin{pmatrix} n_1\gamma_1\zeta \\ +n_2\eta_1\beta \\ +n_3\kappa_1\theta \end{pmatrix}+e\right)}=\frac{n_1\gamma_1\frac{1}{\tau_\zeta}+n_2\eta_1\frac{1}{\tau_\beta}+n_3\kappa_1\frac{1}{\tau_\theta}}{(n_1\gamma_1)^2\frac{1}{\tau_\zeta}+(n_2\eta_1)^2\frac{1}{\tau_\beta}+(n_3\kappa_1)^2\frac{1}{\tau_\theta}+\frac{1}{\tau_e}}.$$

We have four equations to solve for four variables $(\lambda_1,\gamma_1,\eta_1,\kappa_1)$:

$$\gamma_1=\frac{1}{\lambda_1(n_1+1)},\eta_1=\frac{1}{\lambda_1(n_2+1)},\kappa_1=\frac{1}{\lambda_1(n_3+1)},$$

$$\lambda_1=\frac{n_1\gamma_1\frac{1}{\tau_\zeta}+n_2\eta_1\frac{1}{\tau_\beta}+n_3\kappa_1\frac{1}{\tau_\theta}}{(n_1\gamma_1)^2\frac{1}{\tau_\zeta}+(n_2\eta_1)^2\frac{1}{\tau_\beta}+(n_3\kappa_1)^2\frac{1}{\tau_\theta}+\frac{1}{\tau_e}};$$

which gives

$$\lambda_1=\left[\tau_e\left(\frac{n_1}{(n_1+1)^2}\frac{1}{\tau_\zeta}+\frac{n_2}{(n_2+1)^2}\frac{1}{\tau_\beta}+\frac{n_3}{(n_3+1)^2}\frac{1}{\tau_\theta}\right)\right]^{\frac{1}{2}}. \tag{A.5}$$

Proof of Lemma 2: The indifference condition implies

$$\mathbb{E}[\pi_2(\beta)]-c_2=0,\ \mathbb{E}[\pi_3(\theta)]-c_3=0,$$

that is,

$$\frac{\tau_\beta^{-1}}{\lambda_1(n_2+1)^2}-c_2=0,\ \frac{\tau_\theta^{-1}}{\lambda_1(n_3+1)^2}-c_3=0, \tag{A.6}$$

where $\lambda_1=\left[\tau_e\left(\frac{n_1}{(n_1+1)^2}\frac{1}{\tau_\zeta}+\frac{n_2}{(n_2+1)^2}\frac{1}{\tau_\beta}+\frac{n_3}{(n_3+1)^2}\frac{1}{\tau_\theta}\right)\right]^{\frac{1}{2}}$ given in (A.5).

We show that both n_2 and n_3 are increasing in ω. First, by Equation (A.6), we can express n_2 as a function of n_3; that is, $n_2 \equiv n_2(n_3, c_2) = \left(\frac{\tau_\theta}{\tau_\beta}\frac{c_3}{c_2}\right)^{1/2}(n_3+1)-1$, which has the properties that $\frac{\partial n_2}{\partial n_3}>0$ and $\frac{\partial n_2}{\partial c_2}<0$. Second, the equation $\frac{\frac{1}{\tau\theta}}{\lambda_1(n_3+1)^2}-c_3=0$ can be rewritten as $F(n_1,n_3)=\left(\frac{1}{\tau_\theta}\frac{1}{c_3}\right)^2$, where

$$F(n_1,n_3)=\left[\lambda_1(n_3+1)^2\right]^2=\left[\tau_e\left(\frac{n_1}{(n_1+1)^2}\frac{1}{\tau_\zeta}+\frac{n_2(n_3)}{(n_2(n_3)+1)^2}\frac{1}{\tau_\beta}+\frac{n_3}{(n_3+1)^2}\frac{1}{\tau_\theta}\right)\right](n_3+1)^4.$$

It is easy to show that $\frac{\partial F}{\partial n_1}<0$, $\frac{\partial F}{\partial n_3}>0$, and $\frac{\partial F}{\partial c_2}>0$. So by the implicit function theorem $\frac{\partial n_3}{\partial n_1}>0$ and $\frac{\partial n_3}{\partial c_2}<0$. When ω increases, n_1 increases and c_2 decreases, so n_3 is increasing in ω. An increase in ω leads to an increase in n_2 resulting from two forces: an increase in n_3 and a decrease in c_2.

Proof of Proposition 1: Based on the results in the proof of Lemma 1, we have the following results:

$$p=\lambda_1\begin{pmatrix} n_1\gamma_1\zeta \\ +n_2\eta_1\beta \ +e \\ +n_3\kappa_1\theta \end{pmatrix}=\lambda_1\begin{pmatrix} n_1\frac{1}{\lambda_1(n_1+1)}\zeta \\ +n_2\frac{1}{\lambda_1(n_2+1)}\beta \ +e \\ +n_3\frac{1}{\lambda_1(n_3+1)}\theta \end{pmatrix}$$

$$=\left(\frac{n_1}{n_1+1}\zeta+\frac{n_2}{n_2+1}\beta+\frac{n_3}{n_3+1}\theta\right)+\lambda_1 e$$

and

$$\mathrm{var}(p)=\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}+\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}.$$

Then,

$$\begin{aligned}\mathrm{var}(v|p)&=\mathrm{var}(v)-\mathrm{var}(p)\\ &=\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\theta}\right)-\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}+\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\right)\\ &=\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}+\frac{1}{n_3+1}\frac{1}{\tau_\theta}.\end{aligned}$$

Because an increase in ω leads to an increase in n_1, n_2, and n_3, it decreases $\mathrm{var}(v|p)$.

Proof of Proposition 2: The firm manager's investment decision is given by

$$K^*=\mathbb{E}(v|p,\chi,s)=b_1p+b_2\chi+b_3s,$$

where $v=\zeta+\beta+\theta$ and $\begin{pmatrix} p\\ \chi\\ s\end{pmatrix}=\begin{pmatrix}\left(\frac{n_1}{n_1+1}\zeta+\frac{n_2}{n_2+1}\beta+\frac{n_3}{n_3+1}\theta\right)+\lambda_1 e\\ \zeta+\beta+\varepsilon_\chi\\ \theta+\varepsilon_s\end{pmatrix}$ with $e\sim N(0,\tau_e^{-1})$, $\varepsilon_\chi\sim N(0,\tau_\chi^{-1})$ and $\varepsilon_s\sim N(0,\tau_s^{-1})$. So we have

$$\begin{pmatrix} b_1\\ b_2\\ b_3\end{pmatrix}'=\mathrm{cov}\left(v,\begin{pmatrix} p\\ \chi\\ s\end{pmatrix}^T\right)\left(\mathrm{var}\begin{pmatrix} p\\ \chi\\ s\end{pmatrix}\right)^{-1}=\begin{pmatrix}\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}+\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\\ \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\\ \frac{1}{\tau_\theta}\end{pmatrix}^T$$

$$\begin{pmatrix}\left(\frac{n_1}{n_1+1}\right)\frac{1}{\tau_\zeta}+\left(\frac{n_2}{n_2+1}\right)\frac{1}{\tau_\beta}+\left(\frac{n_3}{n_3+1}\right)\frac{1}{\tau_\theta} & \frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta} & \frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\\ \frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta} & \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi} & 0\\ \frac{n_3}{n_3+1}\frac{1}{\tau_\theta} & 0 & \frac{1}{\tau_\theta}+\frac{1}{\tau_s}\end{pmatrix}^{-1}.$$

By calculating the inverse matrix above, it readily follows that

$$b_1=\frac{\begin{pmatrix}\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}+\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\\ \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\\ \frac{1}{\tau_\theta}\end{pmatrix}^T\begin{pmatrix}\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right)\left(\frac{1}{\tau_\theta}+\frac{1}{\tau_s}\right)\\ -\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{\tau_\theta}+\frac{1}{\tau_s}\right)\\ -\left(\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\right)\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right)\end{pmatrix}}{\begin{pmatrix}\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}+\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\\ \frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\\ \frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\end{pmatrix}^T\begin{pmatrix}\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right)\left(\frac{1}{\tau_\theta}+\frac{1}{\tau_s}\right)\\ -\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{\tau_\theta}+\frac{1}{\tau_s}\right)\\ -\left(\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\right)\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right)\end{pmatrix}}=\frac{A}{A+B},$$

where

$$A=\left(\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\right)\left(\frac{1}{\tau_\chi}+\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\right)\left(\frac{1}{\tau_s}\right)+\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{\tau_\theta}+\frac{1}{\tau_s}\right)\frac{1}{\tau_\chi}$$

and

$$B=\left\{\begin{array}{c}\left[\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\right)-\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}\right)\right]\\ \left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{\tau_\theta}+\frac{1}{\tau_s}\right)\end{array}\right\}+\left[\left(\frac{n_3}{n_3+1}\frac{1}{\tau_\theta}\right)\left(\frac{1}{n_3+1}\frac{1}{\tau_\theta}\right)\right]\left(\frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right).$$

When ω increases and hence n_1, n_2, and n_3 increase, A increases and B decreases, so b_1 increases.

Proof of Proposition 3: A sufficient condition for b_4 to decrease in ω is that $\frac{\partial n_2}{\partial\omega}/\frac{\partial n_3}{\partial\omega}$ is high enough (that is, n_2 increases faster enough than n_3 when ω increases). For $\frac{\partial n_2}{\partial\omega}/\frac{\partial n_3}{\partial\omega}$ to be high enough, a sufficient condition is that τ_θ is high enough. We first consider the extreme case $\tau_\theta=+\infty$. It follows that

$$\begin{pmatrix}b_1\\ b_2\\ b_4\end{pmatrix}'=cov\left(v,\begin{pmatrix}p\\ \chi\\ \rho\end{pmatrix}^T\right)\left(var\begin{pmatrix}p\\ \chi\\ \rho\end{pmatrix}\right)^{-1},$$

where

$$b_4=\frac{\begin{pmatrix}\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\\ \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\\ \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\end{pmatrix}\begin{pmatrix}-\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\frac{1}{\tau_\chi}\\ -\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\\ \left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right)\end{pmatrix}}{\begin{pmatrix}\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\\ \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}\\ \frac{1}{\tau_\zeta}+\frac{1}{\tau_\beta}+\frac{1}{\tau_\rho}\end{pmatrix}\begin{pmatrix}-\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\frac{1}{\tau_\chi}\\ -\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\\ \left(\frac{n_1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{n_2}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}+\frac{1}{\tau_\chi}\right)\end{pmatrix}}$$

$$=\frac{\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}\right)\frac{1}{\tau_\chi}}{\left(\frac{1}{n_1+1}\frac{1}{\tau_\zeta}+\frac{1}{n_2+1}\frac{1}{\tau_\beta}\right)\left(\frac{1}{\tau_\chi}+\frac{1}{\tau_\rho}\right)+\frac{1}{\tau_\rho}\frac{1}{\tau_\chi}}.$$

When ω increases, n_1 and n_2 increase, so that b_4 decreases. As all functions are continuous, the result that b_4 is decreasing in ω carries over under the sufficient condition that τ_θ is high enough.

Results in Section 1.6: First, consider cross-sectional Prediction 1. By Equation (A.6), $\frac{\frac{1}{\tau_\beta}}{\lambda_1(n_2+1)^2}-c_2=0$, where λ_1 is bounded from below by a positive number because n_1 is bounded.

So when $\tau_\beta \to +\infty$, $n_2 \to 0$, no matter what n_1 (and hence λ_1) and c_2 are. That is, no traders want to acquire information about β even when the information acquisition cost is at its minimum $c_2(\omega=1)$, i.e., $n_2(\omega)=0$ for all ω. In this case, ETF ownership (ω) does not increase stock price informativeness about factor β. Combining the above result with the result in the proof of Proposition 2, it follows that $\frac{\partial b_1\left(\omega;\tau_\beta=\tau_\beta^L\right)}{\partial\omega} > \frac{\partial b_1\left(\omega;\tau_\beta=\tau_\beta^H\right)}{\partial\omega}$ for some $\tau_\beta^H > \tau_\beta^L$.

Second, consider cross-sectional prediction 2. Write n_I as the number of insider managers.

1) The case of $\left(\tau_s=+\infty, \tau_\chi=0\right)$ and $\left(\tau_s=0, \tau_\chi=0\right)$.

In this case, the firm manager is an insider trader with perfect information about θ and no information about $\zeta+\beta$. So $n_I=1$ and (n_2, n_3) is determined by the equations

$$\frac{\frac{1}{\tau_\beta}}{\lambda_1(n_2+1)^2}-c_2=0,\ \frac{\frac{1}{\tau_\theta}}{\lambda_1(n_3+n_I+1)^2}-c_3=0, \tag{A.7}$$

where λ_1 is given by

$$\lambda_1=\left[\tau_e\left(\frac{n_1}{(n_1+1)^2}\frac{1}{\tau_\zeta}+\frac{n_2}{(n_2+1)^2}\frac{1}{\tau_\beta}+\frac{n_3+n_I}{(n_3+n_I+1)^2}\frac{1}{\tau_\theta}\right)\right]^{\frac{1}{2}}. \tag{A.8}$$

As in Lemma 2, both n_2 and n_3 are increasing in ω. Replacing τ_s with $\tau_s=+\infty$ in the proof of Proposition 2, it is easy to show that $\frac{\partial b_1(\omega)}{\partial\omega}>0$ still holds as in the proposition. Moreover, the expected profit from insider trading is positive.

For the case $\left(\tau_s=0, \tau_\chi=0\right)$, the trading-stage equilibrium is identical to that in the absence of managerial trading in the financial market (i.e., $n_I=0$). As for the firm manager's decision, it follows that $\mathbb{E}(v|p)=p$, that is, $b_1=1$ and $\frac{\partial b_1(\omega)}{\partial\omega}=0$. Moreover, the expected profit for the insider manager from insider trading is zero. Overall, we can therefore conclude that $\frac{\partial b_1(\omega;\tau_s=+\infty,\tau_\chi=0)}{\partial\omega} > \frac{\partial b_1(\omega;\tau_s=0,\tau_\chi=0)}{\partial\omega}$ and $\pi_I\left(\tau_s=+\infty,\tau_\chi=0\right) > \pi_I\left(\tau_s=0,\tau_\chi=0\right)$.

2) The case of $\left(\tau_s=\tau_s^H, \tau_\chi\right)$ and $\left(\tau_s=\tau_s^L, \tau_\chi\right)$ for some $\tau_s^H>\tau_s^L$, where τ_χ is positive and small.

Extending the trading game stage to allowing insider trading, the result in Proposition 1 carries over (with $n_I=1$ in Equations (A.7) and (A.8)); specifically, based on the proof of Proposition 2, as all functions are continuous, the result in case 1) regarding $\frac{\partial b_1(\omega)}{\partial\omega}$ holds, that is, $\frac{\partial b_1\left(\omega;\tau_s=\tau_s^H,\tau_\chi\right)}{\partial\omega} > \frac{\partial b_1\left(\omega;\tau_s=\tau_s^L,\tau_\chi\right)}{\partial\omega}$ for some $\tau_s^H>\tau_s^L$ when τ_χ is small enough. Moreover, $\pi_I\left(\tau_s=\tau_s^H,\tau_\chi\right) > \pi_I\left(\tau_s=\tau_s^L,\tau_\chi\right)$ for some $\tau_s^H>\tau_s^L$.

3) The case of $\left(\tau_s=\tau_s^H, {}_\chi=\tau_\chi^H\right)$ and $\left(\tau_s=\tau_s^L, \tau_\chi=\tau_\chi^L\right)$ for some $\tau_s^H>\tau_s^L$, where τ_χ^H and τ_χ^L are positive and small and $\tau_\chi^H-\tau_\chi^L\geq 0$ is not large.

Based on the proof of Proposition 2, all functions are continuous, and the result in case 2) regarding $\frac{\partial b_1(\omega)}{\partial\omega}$ carries over, that is, $\frac{\partial b_1\left(\omega;\tau_s=\tau_s^H,\tau_\chi=\tau_\chi^H\right)}{\partial\omega} > \frac{\partial b_1\left(\omega;\tau_s=\tau_s^L,\tau_\chi=\tau_\chi^L\right)}{\partial\omega}$ for some $\tau_s^H>\tau_s^L$ when τ_χ^H and τ_χ^L are small enough and $\tau_\chi^H-\tau_\chi^L\geq 0$ is not large. Moreover, $\pi_I\left(\tau_s=\tau_s^H,\tau_\chi=\tau_\chi^H\right) > \pi_I\left(\tau_s=\tau_s^L,\tau_\chi=\tau_\chi^L\right)$ for some $\tau_s^H>\tau_s^L$.

A.2. The ETF Market

This section further motivates the first part of Assumption 1 by endogenizing the ETF market and thereby obtaining a link between ETF ownership and the number of ζ-informed traders. We argue that ETFs increase the total number of factor-informed traders by attracting additional liquidity or noise traders who subsidize information collection. These additional traders could be uninformed speculators (Black 1986) or agents who want to control their long or short exposure to an industry (e.g., oil corporations). Some of the additional informed traders trade in both ETFs and their constituents. We present the arguments in a stylized manner that preserves the equilibrium of Section 1. Suppose there are L ETFs, each of which is an equal-weighted average of a countably finite number of stocks. Every stock has an exogenous number of ζ-informed traders, denoted by

n', who do not trade the ETFs. An ETF l ($l=,\dots,L$) owns one share in each constituent stock. Every ETF attracts additional liquidity or noise trading in the amount of $z_l \sim N(0,v_z)$. The final value of each ETF is $\zeta+\zeta_l'$, with ζ and ζ_l' being independent and identically distributed. This value is realized at date 1', which is between dates 0 and 1. Further, every ETF has a certain number n_l of informed traders who observe ζ exactly. We term these investors "E-traders." The ETF has a standard one-period Kyle (1985)-market, and the market maker in each ETF does not condition on the order flows in the individual stocks or other ETFs. From Subrahmanyam (1991), the equilibrium slope of the pricing function in each market, denoted by λ_E, is given by

$$\lambda_E = \sqrt{\frac{n_l}{(n_l+1)^2 v_z \tau_\zeta}},$$

and each E-trader makes an expected profit in the ETF that equals

$$\pi_l = \left[(n_l+1)^2 \lambda_E \tau_\zeta\right]^{-1}.$$

We assume that the cost of obtaining information about ζ is c_l, and now discuss the equilibrium level of E-traders. To simplify our analysis we assume that the entry condition for these traders in each ETF is based only on the expected profits in the ETF. This can be justified by assuming that the variance of noise trading is very large in the ETF compared to the individual stocks, making the latter a very small part of the expected profit. Thus, we have $\pi_l = c_l$ in equilibrium. Substituting for λ_E in the expression for π_l, we find that the latter increases in $v_z > 0$ and decreases in n_l. Thus, each ETF supports an equilibrium level $n_l^* > 0$ of ζ-informed traders. The total number of E-traders is then Ln_l^*. Now, we discuss how many E-traders would participate in any stock i. Let n_i be the number of E-informed traders in stock i, and let L_i be the number of ETFs holding stock i. We assume that some E-traders that trade an ETF l participate in both the ETF and its constituent securities (the "discretionary" traders).[52] Specifically, let the fraction of discretionary traders in an ETF be ρ. Then, we have that the number of ζ-informed traders in stock i $n_i = \rho n_l^* L_i + n'$, which is increasing in ETF ownership L_i, justifying Assumption 1. Alternatively, suppose that each ETF is a weighted average of several stocks, with stock i receiving a weight of w_i, and that each ETF owns w_i shares in stock i. Now suppose that a subset $n_i' < n_l^*$ of E-traders only trade stock i if the weight w_i exceeds a threshold w_c.[53] Then $\forall w_i > w_c$, $n_i = n_l^* + n'$ whereas $\forall w_i \leq w_c$, n_i is lower at $n_l^* - n_l' + n'$. Since ETF ownership Lw_i is also higher for $w_i > w_c$ than for $w_i \leq w_c$, the effect of the weight is also to cause the number of ζ-informed traders in a stock to increase in ETF ownership of the stock.

Appendix B: Variable Definitions

Analyst coverage: The number of unique analysts who issued earnings forecasts for a firm within a fiscal year. Data have been obtained from I/B/E/S Summary files

1/ASSET: The reciprocal of total assets at the fiscal year-end

Δ***ASSET:*** The annual percentage change in total assets

52 This assumption is in the spirit of Subrahmanyam (1991). He models discretionary liquidity traders as being able to choose between the basket and the constituent portfolio, whereas here we model some informed traders as having the discretion to trade both the ETF and its constituent basket. This behavior can be motivated as facilitating occasional and unmodeled arbitrage between the ETF and its underlying portfolio (Ben-David, Franzoni, and Moussawi 2017).

53 This may be because fixed participation costs in low-weight stocks (Orosel 1998) preclude holding these companies. This threshold can be made continuous by assuming that n_i' is a continuous function of w_i.

BETA: The market beta of the firm's equity over a year, obtained by regressing daily excess stock returns on the excess return of the value-weighted CRSP index

B/M: The book-to-market ratio of a firm's stock, defined as in Fama and French (1992)

CAPX: Capital expenditures at fiscal year-end divided by the beginning-of-year total assets

CAPXRND: The sum of capital expenditures and R&D expenses at the fiscal year-end divided by the beginning-of-year total assets

Cash flow beta: The cash flow beta is obtained by regressing an individual firm's quarterly return on equity (ROE) on industry and market-level return on equity (which we obtain by taking a value-weighted average of firm-specific ROEs by two-digit SIC code, and at the market level, respectively), using the past 5 years of quarterly data (with a minimum of eight observations in the regression). We use the coefficient for industry-level ROE as the firm's cash flow beta

CASH: The ratio of cash and cash equivalent to total assets

Δ***CDS:*** Annual change in the credit default spread at the firm level

CF: Net income before extraordinary items plus depreciation and amortization expenses scaled by lagged total assets

DebtDelayCon: A text-based measure of financing constraints in the debt market following Hoberg and Maksimovic (2015)

DelayCon: A text-based measure of financing constraints following Hoberg and Maksimovic (2015), with higher scores indicating more financially constrained firms

DebtIssue: The amount of debt issuance divided by the beginning-of-year total assets

Earn_com and Earn_firm: We estimate these components, per Bhojraj, Mohanram, and Zhang (2020), who decompose unexpected earnings into common (macro and industry) as well as firm-specific components. As these components are estimated by industry and year, there is no cross-sectional variation in the macro component. So first we estimate $Earn_{it}$, which is annual change in earnings per share excluding extraordinary items (EPSPX) of firm i at the end of fiscal year t scaled by the stock price at the end of fiscal year $t-1$. Then, the common component in our case is the equally weighted average of $Earn_{it}$, by two-digit SIC code ($Earn\_Com_{it}$). $Earn\_Firm_{it}$ is the difference between $Earn_{it}$ and $Earn\_Com_{it}$.

E-index: Constructed by adding one index point for each of the six provisions listed in Bebchuk, Cohen, and Ferrell (2009). A higher index value implies weaker governance

EquityDelayCon: A text-based measure of financing constraints in the equity market following Hoberg and Maksimovic (2015)

EquityIssue: The amount of equity issuance divided by the beginning-of-year total assets

ETF: Firm ownership by U.S. equity ETFs at the fiscal year-end

MktETF: Firm ownership by U.S. equity market ETFs at the fiscal year-end. Market ETFs include those ETFs that physically track broad market indexes, including S&P 500, S&P 1500, Russell 1000, Russell 3000, and NYSE/ NASDAQ Composite Index

NonMktETF: Firm ownership by U.S. equity ETFs not classified as Market ETFs at the fiscal year-end

ETFRET: The equally weighted average return across all ETFs owning shares in the stock

Forecast dispersion: Following Diether, Malloy, and Scherbina (2002), analysts' forecast dispersion is the standard deviation of annual earnings-per-share forecasts scaled by the absolute value of the average outstanding forecast

G-index: Constructed by adding one index point for each of the 24 provisions listed in Gompers, Ishii, and Metrick (2003). A higher index value implies weaker governance

GP: Revenues minus cost of goods sold, scaled by the beginning-of-year total assets

GP_com: The common component of firm gross profitability (GP)

GP_firm: The firm-specific component of firm gross profitability (GP

InsiderProfit: The profitability of insiders' trades over the past 3 years, measured by the 1-month market-adjusted return in absolute value following the directional transaction of the insider. Insider trades include any open market stock transaction initiated by the top-five executives of a firm, obtained from the Thomson Reuters' Insider Filing database

INST: Institutional ownership in the firm at the end of the fiscal year. It is defined as the sum of shares held by institutions from 13F filings at the end of the fiscal year divided by total shares outstanding

INSTR: Residual institutional ownership orthogonalized with respect to ETF ownership using the following annual cross-sectional regressions:

$$INST_{it} = \alpha_0 + \beta_1 ETF_{it} + e_{it}.$$

IVOL: Idiosyncratic volatility

LEV: The sum of long-term and current liabilities scaled by total assets

M&A: Mergers and acquisitions (AQC in Compustat) scaled by beginning-of-year total assets

MFflow: Following Edmans, Goldstein, and Jiang (2012) and Dessaint et al. (2021), we calculate *MFflow* as fund outflow expressed as a percentage of a stock's total dollar trading volume within a quarter, where the price at the previous quarter end is used to compute dollar trading volume. We then take the sum of the quarterly measure within each stock-year

PQ**:** For each firm, *PQ* is the average Tobin's q of its product market peers. We use the Text-based Network Industry Classification (TNIC) to identify peer firms, as developed by Hoberg and Phillips (2010, 2016) using textual analysis of business descriptions of firms in 10-K filings

Sys_PQ and Firm_PQ: We regress a firm's *PQ* on the aggregate market and industry *PQ* (defined at the two-digit SIC level) by using its historical data. *Sys_PQ* is the predicted component from the regression. *Firm_PQ* is the residual component from the regression

PROFVOL: The standard deviation of the annual industry-level profit margin over a 10-year rolling window. Industry-level profit margin is measured as income before extraordinary items divided by the sales of all firms within the same industry, where the industry is defined at the two-digit SIC code level

Q: The market value of equity plus the book value of assets minus the book value of equity scaled by total assets. The book value of equity follows the definition of Fama and French (1992)

Q_fundamental and Q_noise: We regress a firm's Tobin's q on its *MFflow* and control for firm and year fixed effects. *Q_fundamental* is the residual component from the regression, while *Q_noise* is the predicted component from the regression

Sys_Q and Firm_Q: We regress a firm's Tobin's q on the aggregate market and industry Tobin's q (defined at the two-digit SIC level) by using its historical data. *Sys_Q* is the predicted component from the regression. *Firm_Q* is the residual component from the regression

REV: Total revenues scaled by the beginning-of-year total assets

REV_com: The common component of firms' scaled revenue (REV), calculated similarly to the decomposition for earnings

REV_firm: The firm-specific component of firms' scaled revenue (REV), calculated similarly to the decomposition for earnings

$\boldsymbol{RET_{it+3}}$**:** The annualized stock return of firm i over next 3 years from the beginning of fiscal year $t+1$. We require a stock to have at least 1 year of future returns to construct this variable

RND: R&D expenses at the end of fiscal year divided by the beginning-of-year total assets. Missing values are set to zero

ROA: Income before extraordinary items (Compustat item IB) scaled by total assets

SG: Annual growth rate in sales revenue at the firm level

ln(SIZE): The natural logarithm of the firm's market capitalization at the fiscal year-end

SUE, Systematic SUE,* and *Firm SUE: We define *SUE* as the change in split-adjusted quarterly earnings per share from its value four quarters ago, divided by the standard deviation of this change in quarterly earnings over the prior eight quarters (with a minimum requirement of six quarters). We winsorize *SUE* at the 1st and 99th percentiles, and calculate the value-weighted average *SUE* for every industry j (two-digit SIC) in each fiscal quarter t, in order to obtain industry $SUE_{j,t,}$ We apply a similar procedure to obtain the market-level *SUE* at each quarter t, SUE_t. Then we estimate for each firm i the following rolling-window regression, using 6 years of quarterly data, with a minimum requirement of 4 years:

$$SUE_{i,j,t} = a + b * SUE_{j,t} + c * SUE_t + \epsilon_{i,j,t}$$

For each firm's *SUE*, we calculate its *Systematic SUE* as $\hat{c} * SUE_t + \hat{b} * SUE_{j,t}$, where $\hat{b}$ and $\hat{c}$ are the coefficient estimates from the above regression using data up to year $t-1$, respectively. For this computation, and to mitigate the impact of outliers, we winsorize the coefficients $\hat{b}$ and $\hat{c}$ at the 5th and 95th percentiles. *Firm SUE* is then calculated as *SUE – Systematic SUE*

References

Appel, I., T. Gormley, and D. Keim. 2016. Passive investors, not passive owners. *Journal of Financial Economics* 121:111–41.

Baker, M., J. Stein, and J. Wurgler. 2003. When does the market matter? Stock prices and the investment of equity-dependent firms. *Quarterly Journal of Economics* 118:969–1005.

Bakke, T., and T. Whited. 2010. Which firms follow the market? An analysis of corporate investment decisions. *Review of Financial Studies* 23:1941–80.

Bebchuk L, A. Cohen, A. Ferrell. 2009. What matters in corporate governance? *Review of Financial Studies* 22:783–827.

Bebchuk, L., A. Cohen, and S. Hirst. 2017. The agency problems of institutional investors. *Journal of Economic Perspectives* 31:89–102.

Ben-David, I., F. Franzoni, and R. Moussawi. 2018. Do ETFs increase volatility? *Journal of Finance* 73:2471–535.

———. 2017. Exchange traded funds (ETFs). *Annual Review of Financial Economics* 9:169–89.

Benhabib, J., X. Liu, and P. Wang. 2019. Financial markets, the real economy, and self-fulfilling uncertainties. *Journal of Finance* 74:1503–57.

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

Bhojraj, S., P. Mohanram, and S. Zhang. 2020. ETFs and information transfer across firms. *Journal of Accounting and Economics* 70:101336.

Black, F. 1986. Noise, *Journal of Finance* 41:528–43.

Boehmer, B., and E. Boehmer. 2003. Trading your neighbor's ETFs: Competition or fragmentation? *Journal of Banking and Finance* 27:1667–703.

Boehmer, E., and E. Kelley. 2009. Institutional investors and the informational efficiency of prices. *Review of Financial Studies* 22:3563–94.

Bond, P., A. Edmans, and I. Goldstein. 2012. The real effects of financial markets. *Annual Review of Financial Economics* 4:339–60.

Brogaard, J., M. Ringgenberg, and D. Sovich. 2019. The economic impact of index investing. *Review of Financial Studies* 32:3461–99.

Bushee, B. 1998. The influence of institutional investors on myopic R&D investment behavior. *Accounting Review* 73:305–33.

Chen, Q., I. Goldstein, and W. Jiang. 2007. Price informativeness and investment sensitivity to stock price. *Review of Financial Studies* 20:619–50.

Chen, Y., B. Kelly, and W. Wu. 2020. Sophisticated investors and market efficiency: Evidence from a natural experiment. *Journal of Financial Economics* 138:316–41.

Cong, L., and D. Xu. 2019. Rise of factor investing: Asset prices, informational efficiency, and security design. Working Paper, University of Chicago.

Dannhauser, C. 2017. The impact of innovation: Evidence from corporate bond ETFs. *Journal of Financial Economics* 125:537–60.

Da, Z. and S. Shive. 2018. Exchange traded funds and asset return correlation. *European Financial Management* 24:136–68.

Dessaint, O., T. Foucault, L. Frésard, and A. Matray. 2019. Noisy stock prices and corporate investment. *Review of Financial Studies* 32:2625–72.

Dessaint, O., J. Olivier, C. Otto, and D. Thesmar. 2021. CAPM-based company (mis)valuations. *Review of Financial Studies* 34:1–66.

Diether, K., C. Malloy, and A. Scherbina. 2002. Differences of opinion and the cross section of stock returns. *Journal of Finance* 57:2113–41.

Edmans, A., I. Goldstein, and W. Jiang. 2012. The real effects of financial markets: The impact of prices on takeovers. *Journal of Finance* 67:933–71.

Edmans, A., S. Jayaraman, and J. Schneemeier. 2017. The source of information in prices and investment-price sensitivity. *Journal of Financial Economics* 126:74–96.

Ernst, T. 2019. Stock-specific price discovery from ETFs. Working Paper, Massachusetts Institute of Technology.

Fazzari, S., G. Hubbard, and B. Petersen. 1988. Financing constraints and corporate investment. *Brookings Papers on Economic Activity* 1988:141–206.

Ferreira, M., and P. Matos. 2008. The colors of investors' money: The role of institutional investors around the world. *Journal of Financial Economics* 88:499–533.

Foucault, T., and L. Frésard. 2012. Cross-listing, investment sensitivity to stock price, and the learning hypothesis. *Review of Financial Studies* 25:3305–50.

———. 2014. Learning from peers' stock prices and corporate investment. *Journal of Financial Economics* 111:554–77.

———. 2019. Corporate strategy, conformism, and the stock market. *Review of Financial Studies* 32:905–50.

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

Goldstein, I., E. Ozdenoren, and K. Yuan. 2013. Trading frenzies and their impact on real investment. *Journal of Financial Economics* 109:566-82.

Goldstein, I., and L. Yang. 2015. Information diversity and complementarities in trading and information acquisition. *Journal of Finance* 70:1723–65.

———. 2019. Good disclosure, bad disclosure. *Journal of Financial Economics* 131:118–38.

Gompers, P., J. Ishii, and A. Metrick. 2003. Corporate governance and equity prices. *Quarterly Journal of Economics* 118:107–55.

Hirshleifer, D. and S. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36:337–86.

Hoberg, G., and V. Maksimovic. 2015. Redefining financial constraints: A text-based analysis. *Review of Financial Studies* 28:1312–52.

Hoberg, G., and G. Phillips. 2010. Product market synergies and competition in mergers and acquisitions: A text-based analysis. *Review of Financial Studies* 23:3773-811.

———. 2016. Text-based network industries and endogenous product differentiation. *Journal of Political Economy* 124:1423–65.

Hong, H., W. Torous, and R. Valkanov. 2007. Do industries lead stock markets? *Journal of Financial Economics* 83:367–96.

Huang, S., M. O'Hara, and Z. Zhong. 2021. Innovation and informed trading: Evidence from industry ETFs. *Review of Financial Studies* 34:1280–316.

Hwang, B., B. Liu, and W. Xu. 2019. Arbitrage involvement and security prices. *Management Science* 65:2858–75.

Israeli, D., C. Lee, and S. Sridharan. 2017. Is there a dark side to exchange traded funds (ETFs)? An information perspective. *Review of Accounting Studies* 22:1048–83.

Jackson, A., M. Plumlee, and B. Rountree. 2018. Decomposing the market, industry, and firm components of profitability: Implications for forecasts of profitability. *Review of Accounting Studies* 23:1071–95.

Jayaraman, S., and J. Wu. 2019. Is silence golden? Real effects of mandatory disclosure. *Review of Financial Studies* 32:2225–59.

John, K., L. Litov, and B. Yeung. 2008. Corporate Governance and Risk-Taking. *Journal of Finance* 63:1679-1728.

Kahraman, B. 2021. Publicizing arbitrage. *Journal of Financial and Quantitative Analysis* 56:789–820.

Lee, C., and E. Watts. 2021. Tick size tolls: Can a trading slowdown improve earnings news discovery? *Accounting Review* 96:373–401.

Li, F. W., and Q. Zhu. 2022. Short selling ETFs. *Review of Asset Pricing Studies*. Advance Access published February 3, 2022, 10.1093/rapstu/raac005.

Lin, T., Q. Liu, and B. Sun. 2019. Contractual managerial incentives with stock price feedback. *American Economic Review* 109:2446–68.

Luo, Y. 2005. Do insiders learn from outsiders? Evidence from mergers and acquisitions. *Journal of Finance* 60:1951–81.

Malkiel, B. G. 2016. Is indexing worse than Marxism? *Wall Street Journal*, November 24. http://tinyurl.com/5n7xp55e.

Orosel, G. 1998. Participation costs, trend chasing, and volatility of stock prices. *Review of Financial Studies* 11:521–57.

Panousi, V., and D. Papanikolaou. 2012. Investment, idiosyncratic risk, and ownership. *Journal of Finance* 67:1113–48.

Peters, R., and L. Taylor. 2017. Intangible capital and the investment-q relation. *Journal of Financial Economics* 123:251–72.

Polk, C., and P. Sapienza. 2009. The stock market and corporate investment: A test of catering theory. *Review of Financial Studies* 22:187–217.

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

Sockin, M., and W. Xiong. 2015. Informational frictions and commodity markets. *Journal of Finance* 70:2063–98.

Subrahmanyam, A. 1991. A theory of trading in stock index futures. *Review of Financial Studies* 4:17–51.

Subrahmanyam, A., and S. Titman. 1999. The going-public decision and the development of financial markets. *Journal of Finance* 54:1045–82.

Wardlaw, M. 2020. Measuring mutual fund flow pressure as shock to stock returns. *Journal of Finance* 75:3221–43.

Yan, D. 2017. Do private firms (mis) learn from the stock market? Working Paper, Stockholm School of Economics.

Zou, Y. 2019. Lost in the rising tide: ETF flows and valuation. Working Paper, Harvard University.

Zuo, L. 2016. The informational feedback effect of stock prices on management forecasts. *Journal of Accounting and Economics* 61:391–413.