Pacific-Basin Finance Journal 79 (2023) 102005

Contents lists available at ScienceDirect

Pacific-Basin Finance Journal

journal homepage: www.elsevier.com/locate/pacfin

ETF ownership and informational efficiency of underlying stocks: Evidence from China☆

Weili Wu, Feifei Zhu *

School of Finance, Central University of Finance and Economics, China

ARTICLE INFO

JEL classification codes:
G12
G14
M41

Keywords:
Exchange-traded funds (ETF)
Informational efficiency
Stock liquidity
Informed trading

ABSTRACT

This study investigates how exchange-traded fund (ETF) activity affects the informational efficiency of its underlying stocks in the Chinese market, which has several different features from developed markets. We show that increased ETF ownership improves stock liquidity and attracts informed investors, resulting in higher pricing efficiency. By exploiting the heterogeneity of four different types of ETFs in the Chinese market, we show that the informational role of ETFs depends on whether they can be used for intraday trading, which significantly improves underlying securities' liquidity. Our study adds to the ongoing debate on the economic consequences of financial innovations and sheds light on the informational role of ETFs for emerging markets.

1. Introduction

Financial innovations during the last two decades have produced a bewildering array of composite securities. Examples of such securities baskets include a large variety of mortgage- and asset-backed securities, index-linked securities, and real estate investment trusts. Among these, exchange-traded funds (ETFs) have witnessed tremendous growth. Since the inception of the first ETF (the Standard and Poor's Depository Receipts) in 1993, there has been exponential growth of ETF products in the global market. Their total market capitalization climbed to approximately $8 trillion by the end of 2020, with a compound annual growth rate (CAGR) of 18.4% over the past decade. In China, the first ETF (Huaxia SSE 50ETF) was launched on February 23, 2005. By the end of 2020, China had 371 ETF products, with assets under management reaching nearly 1.1 trillion RMB and a CAGR of 31.2% over the past decade.

Given the rapid development of the ETF market, the literature on the economic consequences of ETFs remains nascent (Lettau and Madhavan, 2018; Glosten et al., 2021). The Securities and Exchange Commission (SEC) has called for more research and discussion on the economic consequences of ETFs, especially after the "flash crash" on May 6, 2010.[1] In addition, early evidence shows a growing debate on the economic effects of index-linked products on the stock market (Lettau and Madhavan, 2018). Some studies suggest that ETFs distort capital markets, increase stock volatility, and contribute to return comovement (Da and Shive, 2018; Ben David et al., 2018), whereas others highlight the positive effects of ETFs in price discovery and improving market quality (Huang et al., 2021;

☆ Usual disclaimers apply. All errors are our own. We would like to thank Wen Chen, He Ni, Yunhong Yang and seminar participants at Conference on Empirical Research of Accounting and Finance Issues in China, China Accounting and Finance Conference. We acknowledges financial support from the National Science Foundation of China (grant number 72102247, 71702205, 72273160), and the Program for Innovation Research in Central University of Finance and Economics..

* Corresponding author at: School of Finance, Central University of Finance and Economics, Changping District, Beijing 10026, China.
E-mail address: feifei.zhu@cufe.edu.cn (F. Zhu).

[1] For findings regarding the market events, the SEC report can be accessed at https://www.sec.gov/files/marketevents-report.pdf.

https://doi.org/10.1016/j.pacfin.2023.102005
Received 25 September 2022; Received in revised form 7 March 2023; Accepted 8 March 2023
Available online 11 March 2023

Glosten et al., 2021; Antoniou et al., 2020).

In this study, we investigate how ETF activity affects the informational efficiency of its underlying stocks in the Chinese A-share market. As surveyed by Bond et al. (2012), the information quality of an individual security has important implications for firms' investment decision efficiency and capital allocation. Thus, our research question is of considerable interest to regulators, practitioners, and researchers alike. In addition, extant studies find contrasting results about the effect of ETF ownership on the informational efficiency of component securities. Some studies show that ETFs improve informational efficiency for the underlying stocks in the short run (Glosten et al., 2021), whereas others note a negative relation between ETF ownership and the pricing efficiency of underlying stocks (Israeli et al., 2017). Thus, the jury is still out on how an increase in ETF ownership affects the pricing efficiency of underlying stocks, and further in-depth study is needed.

The Chinese A-share market offers several interesting perspectives and thus deserves further study (Carpenter et al., 2021). For one, it has been the second-largest stock market in the world regarding market capitalization, with a record high of more than \$10 trillion. The rapid growth of this market has fueled a growing body of literature on this market in financial economics. In addition, the special features of this market, such as the "T + 1" trading restriction, 10% daily price limits, dominance of retail investors, and short-sale constraints, distinguish it from developed markets. Therefore, the role ETFs play in this market may differ from their role in other markets, which is worth studying.

To address this question, we first conduct a series of analyses using a panel of stock-year observations between 2006 and 2020. Following prior literature, we construct five proxies for stock price informational efficiency: non-synchronicity of stock price (*Nonsynch*), the dynamic relation between returns and volumes based on a regression analysis of difference or residual terms ($LMSW_{dif}$ and $LMSW_{res}$), and variance ratio (VR) based on 6- or 12-period lags [$VR(6)$ and $VR(12)$]. The empirical results show that increases in ETF ownership are associated with increases in stock price informativeness. For instance, a one percentage point increase in ETF ownership is associated with a 12.93 percentage point increase in *Nonsynch* over the next year.

In addition, we explore the relation between the changes in ETF ownership and stock pricing efficiency by examining the influence of ETF activities on the sensitivity of stock returns to their future earnings. The results indicate that ETF activities facilitate the incorporation of future earnings information into underlying stock prices and thus improve the informational efficiency of the underlying securities. Furthermore, to examine whether such an improvement should be attributed to firm-specific or macro-based information, or both, we decompose total earnings into "macro-based" and "firm-specific" components. Our results show that interactions of ETF ownership and firm-specific earnings components are significantly positive, which indicates that the contribution of ETF activity to pricing efficiency can be attributable to its pricing discovery in firm-specific information.

To solve the endogeneity problem, we employ CSI 300 and CSI 500 index reconstitution as an exogenous shock to ETF ownership and exploit the instrumental variable (IV) method. CSI 300/500 index reconstitution offers a setting in which firms with similar characteristics have a significant variation in ETF ownership, while it has no direct effect on stock pricing efficiency. Our first-stage regression verifies that a firm transitioning from the CSI 300 to the CSI 500 experiences an increase in ETF ownership relative to a firm transitioning from the CSI 500 to the CSI 300, as the indexes are value-weighted. In the second-stage regression, our results indicate that stocks that experienced a significant increase in ETF ownership due to index reconstitution have higher informational efficiency. The above results corroborate our main findings.

Regarding the underlying channels, drawing the insight from the price discovery literature (Hasbrouck, 2003; Yu, 2005; Chen and Strother, 2008; Ivanov et al., 2013), we examine the relation between ETF ownership, stock liquidity, informed trading, and the pricing efficiency of underlying stocks. Our results show that increased ETF ownership leads to higher stock liquidity. Furthermore, higher stock liquidity attracts more informed investors to stock trading. Consequently, both stock liquidity and the participation of informed investors contribute to the pricing efficiency of underlying stocks.

We also exploit the unique features of ETFs in the Chinese market to verify our underlying mechanism. As ETFs are traded in both primary and secondary markets, investors can use them to circumvent the "T + 1" trading restriction in the Chinese market and conduct intraday trading. There are four types of ETFs in the Chinese market: Shanghai single-market ETFs, Shenzhen single-market ETFs, Shanghai cross-market ETFs, and Shenzhen cross-market ETFs. The first three have sufficiently high settlement efficiency; thus, they can be used to conduct intraday trading. However, the latter cannot because of its low settlement efficiency. We exploit this heterogeneity and examine the impacts of different types of ETFs. We find that the informational efficiency of underlying stocks increases with the ownership of all other types of ETFs, except for Shenzhen cross-market ETFs. As Holden (1995) and Subrahmanyam and Titman (1999) suggested, if ETFs cannot be used to conduct intraday arbitrage, their ability to provide liquidity to underlying stocks will be severely weakened. As the Shenzhen cross-market ETFs cannot be used to conduct intraday trading, they have limited ability to provide liquidity and attract informed investors. Therefore, the nonsignificant impact of Shenzhen cross-market ETFs further verifies our results of channel tests.

In addition, we also replicate the empirical analyses in Glosten et al. (2021) and test whether ETF activity would improve short-run informational efficiency. We find that ETF activity increases the return-earnings relation and thus improve the informational efficiency of underlying stocks in the short-run, which is similar to the findings of Glosten et al. (2021). However, different from their results, we find that the increase in short-run informational efficiency in the Chinese A-share market is attributable to contemporaneous firm-specific earning information rather than systematic accounting information.

The contribution of this study is twofold. First, we contribute to the growing debate on the consequences of ETFs. One strand of these studies shows that ETFs increase return comovement (Da and Shive, 2018), exaggerate stock volatility (Ben David et al., 2018), and harm stock liquidity (Israeli et al., 2017). Other studies suggest that trading associated with the ETF-arbitrage mechanism can improve intraday price discovery for the underlying stocks (Hasbrouck, 2003; Yu, 2005; Chen and Strother, 2008; Ivanov et al., 2013). In this study, we use Chinese A-share market data and find positive effects of ETF ownership on the informational efficiency of the underlying stocks, which broadens this strand of literature.

Second, we provide empirical evidence depicting the channels through which ETF activities increase the pricing efficiency of underlying stocks and differentiate our paper from previous studies. Different from Israeli et al. (2017), who find that ETF activity leads to the deterioration of pricing efficiency, we find that increases in ETF ownership are associated with increases in stock price informativeness. The contradicting results lie in the effect of ETFs on stock liquidity in different stock markets. In the context of the Chinese A-share market, which implements the "T + 1" trading restriction, ETFs could increase the liquidity of underlying stocks through intraday trading. We utilize the heterogeneity of four different types of ETFs in the Chinese market and corroborate this idea. In addition, different from Glosten et al. (2021), who find that ETF activity does not predict future fundamentals but merely incorporates systemic earnings news rather than firm-level information in a timely manner. We .find that changes in ETF ownership have a positive effect on stock pricing efficiency in both the long run and short run. This can be attributed to its pricing discovery in firm-specific information rather than systematic information.

The rest of the paper proceeds as follows. Section 2 introduces the institutional background and proposes two competing hypotheses. Section 3 shows our variable construction and descriptive statistics. Section 4 presents our empirical results. Section 5 discusses the heterogeneous effects of the four different types of ETFs and our robustness tests. Section 6 concludes.

2. Institutional background and hypothesis development

2.1. Institutional background

The Chinese A-share market implements the "T + 1" trading restriction in its secondary market; that is, stocks bought today cannot be sold until tomorrow. This restriction is intended to stabilize the market and inhibit excessive trading, but it also stifles the flexibility of investors, hinders timely and adequate reflection of the information in the stock price and is detrimental to risk management. However, by trading ETFs, investors can conduct intraday trading of their constituent stocks and circumvent the "T + 1" trading restriction, which amplifies the advantages of ETFs in the Chinese A-share market. Note that ETFs live in both the primary market, where units of ETF shares can be swapped for pre-announced portfolios of the underlying assets, and in the secondary market, where investors can buy or sell them. This feature allows investors to conduct intraday trade on ETFs and their underlying stocks. Specifically, there are two ways to achieve this goal. First, investors can buy ETF constituent stocks in the secondary market, then use them to subscript ETF shares in the primary market, and finally sell ETF shares in the secondary market to achieve intraday trading. Second, investors can buy ETF shares in the secondary market, then redeem ETF shares in the primary market to obtain ETF constituent shares, and finally sell the constituent shares in the secondary market to perform intraday trading.

Interestingly, among the four types of ETFs in the Chinese market, not all of them can be used for intraday trading. Specifically, based on the listed market, ETFs can be divided into "Shanghai stock exchange ETFs" if they are listed on the Shanghai Stock Exchange (SHSZ) and "Shenzhen stock exchange ETFs" if they are listed on the Shenzhen Stock Exchange (SZSZ). Based on the trading market of their component securities, ETFs can be divided into "single-market ETFs" if all component securities are traded in a single stock exchange and "cross-market ETFs" if component securities include stocks that are traded in either the SHSZ or SZSZ. Thus, ETFs can be grouped into four types: Shanghai single-market ETFs (ETF_{SHS}), Shenzhen single-market ETFs (ETF_{SZS}), Shanghai cross-market ETFs (ETF_{SHC}), and Shenzhen cross-market ETFs (ETF_{SZC}).

Different types of ETFs have different settlement efficiencies, and only ETFs with high settlement efficiencies can be used to conduct intraday trading. Among the four types of ETFs, ETF_{SHS}, ETF_{SZS}, and ETF_{SHC} have high settlement efficiency, but ETF_{SZC} has low settlement efficiency.[2] Consequently, investors can use the former three types of ETFs to conduct intraday trading but cannot use ETF_{SZC} to do so. By exploiting this unique setting, we verify the mechanism through which ETF activities affect the pricing efficiency of underlying stocks, which is detailed in Section 5.1.

2.2. Hypothesis development

There are two contrasting predictions on the possible impacts of ETF ownership on the pricing efficiency of underlying securities. On the one hand, ETFs can facilitate the pricing efficiency of underlying stocks by increasing stock liquidity and attracting more informed investors to trade the underlying stocks (Hasbrouck, 2003; Yu, 2005; Chen and Strother, 2008; Ivanov et al., 2013).

Generally, ETFs could increase the liquidity of underlying stocks in two ways: index arbitraging or intraday trading. In information-based models such as Fremault (1991), Kumar and Seppi (1994), and Holden (1995), index arbitrageurs can exploit the price differences between the stock market and index-linked securities (such as ETFs) and conduct index arbitraging and, therefore, provide

[2] For single-market ETFs, investors can subscript or redeem ETFs, as well as their component securities, through a single stock exchange. Therefore, their settlement efficiency is high, and it is far easier for investors to perform intraday trading. However, for cross-market ETFs, subscription and redemption of ETFs and their component securities are done through different stock exchanges; thus, the settlement efficiency is relatively low, and it is more difficult for investors to conduct intraday trading. Furthermore, based on the listed markets, differences in settlement efficiencies still exist among cross-market ETFs. For the SHSZ, when investors subscript or redeem cross-market ETFs, it allows them to use cash to replace component stocks traded in other exchanges, so investors can still enjoy high settlement efficiency and realize intraday trading. However, because the SESZ adopts a "physical subscription and redemption" mechanism for stocks traded in other exchanges; thus, investors cannot conduct intraday trading by exploiting cross-market ETFs listed in this market.

liquidity to the market (Brennan and Schwartz, 1990). In addition, as ETFs help investors perform intraday trading, investors exploiting intraday trading strategies could also provide liquidity to investors with interday trading strategies, thus improving stock liquidity (Subrahmanyam and Titman, 1999).

Furthermore, as Kyle (1985), Admati and Pfleiderer (1988), and Foster and Viswanathan (1990) suggested, higher stock liquidity can induce informed traders to trade on private information. With more participation of informed investors, the pricing efficiency of underlying stocks would increase accordingly. Thus, higher ETF ownership is accompanied by higher informational efficiency of underlying stocks.

Particularly, in the Chinese A-share market, which features certain trading restrictions (such as the "T + 1" trading mechanism), the positive relation between ETF ownership and pricing informativeness is more salient. For instance, to circumvent the trading restriction, investors would exploit ETFs when conducting intraday trading, which would increase the liquidity of underlying stocks and result in an increase in stock pricing efficiency. Based on the reasoning outlined above, we propose the following hypothesis.

H1a. An increase in ETF ownership is associated with higher informational efficiency of underlying stocks.

In contrast, the relation between ETF ownership and pricing efficiency can be negative. As ETF ownership grows, an increasing proportion of the outstanding shares for the underlying security becomes "locked up" by the fund sponsor, leading to a decrease in stock liquidity. Furthermore, ETFs offer an attractive alternative investment vehicle for uninformed (or "noise") traders, who would otherwise trade underlying component securities and provide liquidity (Kyle, 1985). The noisy rational expectations literature (Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981; Verrecchia, 1982) suggests that when uninformed traders leave the market, informed traders lose their counterparts, so they quit the market as well. Consequently, stock price informativeness will fall as a consequence of the emergence of ETFs.

In sum, as ETF ownership increases, both the locking up of shares and the migration of uninformed investors away from underlying component securities create a steady siphoning of firm-level liquidity, which generates a disincentive for informed traders to acquire information and thus decreases the informational efficiency of underlying securities. Accordingly, we propose the following alternative hypothesis:

H1b. An increase in ETF ownership is associated with lower informational efficiency of underlying stocks.

3. Variable construction and descriptive statistics

3.1. Variable construction

3.1.1. Informational efficiency

3.1.1.1. Stock return non-synchronicity. Following prior literature (Roll, 1984; Wurgler, 2000; Durnev et al., 2003), we exploit $Nonsynch_{it}$ as a proxy for stock price informativeness. It measures the extent to which variation in firm-level stock returns is not attributable to movements in market and related-industry returns. Specifically, we follow the methodology outlined in Durnev et al. (2003). First, for each stock-year observation, we obtain the adjusted coefficient of determination (adjusted R^2) by regressing daily stock returns on the current and prior day's value-weighted market returns and the current and prior day's value-weighted industry returns:

$$R_{id} = \alpha + \beta_1 R_d^M + \beta_2 R_{d-1}^M + \beta_3 R_{i,d}^I + \beta_4 R_{i,d-1}^I + \varepsilon_{id} \tag{1}$$

where R_{id} is stock i's return on day d, R_d^M is the value-weighted market return on day d, and R_{id}^I is the value-weighted return of stock i's industry, defined using the Global Industries Classification Standard (GICS),[3] on day d. Eq. (1) is estimated separately for each stock year, using daily returns of stock i over the trading days in year t, with a minimum of 150 observations.

Next, for each stock-year observation, we calculate the annual measure of stock return non-synchronicity, $Nonsynch_{it}$, as the logarithmic transformation of R_{it}^2 to create an unbounded continuous measure of non-synchronicity:

$$Nonsynch_{it} = log\frac{1 - R_{it}^2}{R_{it}^2} \tag{2}$$

A high value of *Nonsynch* indicates a low fraction of variation of firm-level stock returns, which can be explained by variations in the market and related-industry returns and, therefore, high stock-price efficiency.

3.1.1.2. Dynamic volume–return relation. Our second proxy for stock price informativeness is proposed by Llorente et al. (2002) (hereafter LMSW), who posit that returns generated by risk-sharing trades tend to reverse themselves, whereas returns generated by speculative trades tend to continue themselves. The dynamic relation between return and volume is estimated in the following model:

$$R_{i,d+1} = \beta_0 + \beta_1 R_{id} + \beta_2 V_{id} R_{id} + \varepsilon_{i,d+1} \tag{3}$$

[3] The GICS was developed jointly by Morgan Stanley Capital International (MSCI) and Standard & Poors. The GICS methodology is widely used by the MSCI indexes as well as by a large portion of the professional investment management community.

where R_{id} denotes the return of stock i on day d, and V_{id} is the detrended log turnover. Since the daily time series of turnover is nonstationary, Llorente et al. (2002) measure turnover in logs and detrend the resulting series by subtracting a 200-trading-day moving average. Specifically, V_{id} is defined as $V_{id} = ln(turn_{id}) - \frac{1}{200}\sum_{s=1}^{200} ln(turn_{id-s})$, where $turn_{id}$ represents stock i's daily turnover as the total number of shares traded in that day divided by the total number of shares outstanding.

In principle, trading in every stock contains both hedging and speculative elements. The observed volume–return relation depends on the relative importance of one type of trade relative to the other. Positive β_2 coefficients are expected for stocks that are associated with significant speculative trade, while for stocks with predominantly hedging trade, the β_2 coefficients should be negative. Stocks for which neither speculative nor hedging trade dominates should have β_2 coefficients that are insignificantly different from zero. Thus, the β_2 coefficient represents the level of informational efficiency; a higher coefficient magnitude is associated with higher efficiency of the stock price.

Furthermore, to rule out the possible influences of market-wide information, we use two methods to obtain the firm-specific informational efficiency measures.

The first is to estimate model (4):

$$R_{res,i,d+1} = \beta_0 + \beta_1 R_{res,i,d} + \beta_2 V_{res,i,d} R_{res,i,d} + \varepsilon_{i,d+1} \tag{4}$$

where $R_{res,\ i,\ d}$ is the residual term obtained from regressing stock i's return on day d on the value-weighted market return on day d; and $V_{res,\ i,\ d}$ is the residual term obtained from regressing stock i's detrended log turnover on day d on the value-weighted market-wide detrended log turnover on day d. For the estimated coefficient β_2, we denote it as $LMSW_{res}$ and use it as the proxy for stock price informativeness.

The second method is to estimate model (5):

$$R_{dif,i,d+1} = \beta_0 + \beta_1 R_{dif,i,d} + \beta_2 V_{dif,i,d} R_{dif,i,d} + \varepsilon_{i,d+1} \tag{5}$$

where $R_{dif,\ i,\ d}$ ($V_{dif,\ i,\ d}$) is the difference between stock i's return (detrended log turnover) on day d and the value-weighted market return (market-wide detrended log turnover) on day d. For the estimated coefficient β_2, we denote it as $LMSW_{dif}$ and use it as another firm-specific informational efficiency measure.

For both $LMSW_{res}$ and $LMSW_{dif}$, higher values are associated with higher informational efficiency of stock prices.

3.1.1.3. Variance ratio. Following Lo and MacKinlay (1988), we also exploit the "variance ratio" (*VR*) to measure stock price informational efficiency. If a series follows the random walk process, the variance of its q-order difference (q > 1) will be q times its first-order difference variance. Mathematically,

$$\mathrm{Var}\left(P_t - P_{t-q}\right) = q\mathrm{Var}(P_t - P_{t-1}) \tag{6}$$

where P_t denotes the stock price on day t. The VR test statistic is defined as

$$VR(q) = \left| \frac{\mathrm{Var}\left(P_t - P_{t-q}\right)}{q\mathrm{Var}(P_t - P_{t-1})} - 1 \right| \tag{7}$$

Under the assumption of a random walk process, the statistics should equal zero, with a value greater than zero implying a positive or negative autocorrelation of the disturbance. Thus, lower values of *VR(q)* are accompanied by a higher efficiency of information. We set q equal to 6 and 12, obtain *VR(6)* and *VR(12)*, and use them to evaluate stock price informativeness.

3.1.1.4. Future earning response coefficient. Another proxy for the informational efficiency of stock prices is the future earnings response coefficient (FERC), which measures the extent to which current stock returns reflect future firm earnings (Bond et al., 2012; Israeli et al., 2017; Glosten et al., 2021). A higher FERC is associated with more fundamental earnings information being incorporated into underlying stock prices. The detailed model specification of this test will be illustrated in Section 1.1.6.

3.1.2. ETF activities

We use changes in ETF ownership as the proxy for the ETF activity of each stock because it is a direct measure that aggregates the net activities of ETFs at the stock level. ETF ownership is calculated as the proportion of shares owned by all candidate ETFs in a stock's total shares outstanding. Specifically, the ETF ownership (level variable) for stock i at the end of year t, $ETF_{i,\ t}$, is calculated as

$$ETF_{i,t} = \frac{\sum_{j=1}^{J} Shares_{i,j,t}}{Totol\ share\ Outstanding_{i,t}} \tag{8}$$

where j is the set of ETFs holding stock i, $Shares_{i,\ j,\ t}$ is the number of stock i's shares held by ETF j at the end of year t, and $Totol\ share\ Outstanding_{i,\ t}$ is the total shares outstanding for stock i at the end of year t. The change in ETF ownership is calculated as the annual difference in $ETF_{i,\ t}$.

In addition, for the different types of ETF (i.e., ETF_{SZC}, ETF_{SHC}, ETF_{SZS}, and ETF_{SHS}), we also calculate their corresponding measures (both level and change variables) for each stock in each year. The change variable is the difference between level variables in two consecutive years.

3.1.3. Stock liquidity

We use three proxies for stock liquidity: (1) share turnover (*Turn*), (2) the Amivest measure (*Amivest*), and (3) the Gamma measure (*Gamma*).

As a widely used measure in the asset pricing literature, share turnover (*Turn*) measures how easy it is to sell shares on the open market. The higher the share turnover, the more liquid the stock. We define it as the equal-weighted average of the ratio of daily trading volume to daily total market value, as shown in Eq. (9).

$$Turn_{it} = \frac{1}{N_{it}} \sum_{d=1}^{N_{it}} \frac{volume_{id}}{TMV_{id}} \tag{9}$$

where $volume_{id}$ and TMV_{id} are the trading volume and total market value of stock i on day d, and N_{it} is the number of trading days within one year.

The Amivest measure (*Amivest*) determines the trading volume associated with a unit change in the stock price (Amihud et al., 1997). A higher *Amivest* implies greater market liquidity or depth. *Amivest* is defined as

$$Amivest_{it} = \frac{1}{N_{it}} \sum_{d=1}^{N_{it}} \frac{volume_{id}}{|R_{id}|} \times 10^{-9} \tag{10}$$

where $volume_{id}$ and R_{id} are the trading volume and return of stock i on day d, respectively. The summation is over days in the estimation period (i.e., a year). Following Amihud et al. (1997), we calculate the measure regarding logs to reduce the impact of extreme observations.

Following Pástor and Stambaugh (2003), we also use the Gamma measure (*Gamma*) as a proxy for stock liquidity. The liquidity for stock i in year t is the ordinary-least-square estimate of $\gamma_{i,\ t}$ in the regression analysis:

$$R^e_{i,d+1} = \theta_i + \varphi_i R_{i,d} + \gamma_i sign\left(R^e_{i,d}\right) \cdot dvol_{i,d} + \varepsilon_{i,d+1} \tag{11}$$

where $R^e_{i,\ d}$ represents the excess return of stock i on day d. It equals the raw return of stock i on day d minus the value-weighted market return on day d. $dvol_{i,\ d}$ denotes the dollar volume for stock i on day d. Sign(.) is a sign function that equals positive one when $R^e_{i,\ d}$ is greater than zero, negative one when $R^e_{i,\ d}$ is smaller than zero, and zero when $R^e_{i,\ d}$ equals zero.

The basic idea is that "order flow", constructed here simply as volume signed by the contemporaneous excess return of a stock, should be accompanied by a return that one expects to be partially reversed in the future if the stock is not perfectly liquid. The greater the expected reversal for a given dollar volume, the lower the stock's liquidity. That is, $\gamma_{i,\ t}$ should be negative in general and larger in absolute magnitude when liquidity is lower.

3.1.4. Informed trading

The probability of informed trading (PIN) is a widely used informed trading measure. The original PIN estimation approach was proposed by Easley et al. (1996), which requires the subjective estimation of unobservable parameters and the application of numerical methods. To overcome the difficulties of estimating the PIN model, Easley et al. (2012) propose a new procedure to estimate informed trading based on a process subordinated to volume arrival, which they name Volume-Synchronized Probability of Informed Trading (VPIN). This new measure does not require the intermediate numerical estimation of unobservable parameters and is updated in stochastic time, which is calibrated to have an equal volume of trade in each time interval. Thus, the VPIN methodology overcomes the difficulties of estimating PIN models in highly active markets and provides an analytically tractable way to measure informed trading.

In the Chinese market, studies have found that the PIN metric is not valid (Han et al., 2008), while most of the literature supports the validity of VPIN (Zhou et al., 2015; Chen et al., 2019). Thus, we use the VPIN measure extended by Easley et al. (2012) to estimate informed trading. Specifically, we calculate the buy and sell volumes using one-minute time bars using Eq. (12):

$$\begin{cases} V^B_\tau = \displaystyle\sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right) \\ V^S_\tau = \displaystyle\sum_{i=t(\tau-1)+1}^{t(\tau)} V_i \cdot \left[1 - Z\left(\frac{P_i - P_{i-1}}{\sigma_{\Delta P}}\right)\right] = V - V^B_\tau \end{cases} \tag{12}$$

where $t(\tau)$ is the index of the last time bar included in the τ volume bucket, Z is the CDF of the standard normal distribution, P_i is the price in time i, and $\sigma_{\Delta P}$ is the estimate of the standard derivation of price changes between time bars. If there is no price change from the beginning to the end of the time bar, we split the volume in a time bar equally between buy and sell volume. If the price increases, the volume is weighted more toward buys than sells, and the weighting depends on how large the price change is relative to the distribution of price changes.

Thus, the VPIN metric can be expressed as follows.

$$VPIN = \frac{\sum_{\tau=1}^{n} \left| V_{\tau}^{B} - V_{\tau}^{S} \right|}{nV} \tag{13}$$

where V is the volume in every bucket, and n is the number of buckets used to approximate the expected trade imbalance.

Following Chen et al. (2019) and Easley et al. (2012), we choose $n = 8$ and $n = 50$ to calculate the VPIN matrix over 8 or 50 buckets and construct the corresponding *VPIN(8)* and *VPIN(50)* measures. For the annual measure, we use the arithmetic mean of all daily VPIN values for the year as the proxy.

3.2. Sample construction and descriptive statistics

We retrieve information on all equity-typed ETFs traded on the Chinese A-share market from the China Stock Market & Accounting Research (CSMAR) database. Specifically, we obtain the reported holdings of ETFs on each stock to construct ETF ownership. For some ETFs, the CSMAR database does not provide regular reporting of equity holdings. In these instances, we hand-collect additional holdings from their annual reports.

In addition, we retrieve daily stock returns, market returns, and industry-level returns, as well as annual accounting information for stocks from the CSMAR database. We clean these data along several dimensions: (1) financial firms, firms marked as special treated (ST) or particular transfer (PT), and delisted firms are excluded from the sample; (2) observations with severe missing data on main variables are also excluded; and (3) all the continuous variables are winsorized at the 1% and 99% percentiles to alleviate the impact of outliers.

Our sample period is from 2006 through 2020. It begins in 2006 because the first equity-typed ETF (Huaxia SSE 50ETF) was listed on February 23, 2005. We exploit the GICS to classify industries, data for which are obtained from the Wind database. The above sample selection process results in 28,724 firm-year observations, with 3610 listed firms and 589 ETFs covered.

Table 1 presents the descriptive statistics for our main variables. Of particular interest for our analyses are the levels and changes in ETF ownership, respectively measured as the percentage of total shares outstanding held by all ETFs and the difference in levels of ETF ownership between two consecutive years. The mean (median) percentage of changes in ETF ownership is 0.22% (0.01%). For the level variable, the mean percentage of ETF ownership is 1.96%, which consists of 0.22% Shenzhen cross-market ETFs (ETF_{SZC}), 0.94% Shanghai cross-market ETFs (ETF_{SHC}), 0.47% Shenzhen single-market ETFs (ETF_{SZS}), and 0.33% Shanghai single-market ETFs (ETF_{SHS}).

Regarding stock price informational efficiency, the mean and median values of changes in *Nonsynch* are both 0.05, with the mean and median levels of *Nonsynch* being 0.39 and 0.36, respectively. The 25-percentile value of Δ*Nonsynch* is lower than zero, which means that the informational efficiency of at least 25% of observations deteriorates during two consecutive years. For $\Delta LMSW_{res}$ and ΔLMSW_{dif}, the mean and median values all equal zero, displaying normal distributions of changes in dynamic volume–return relations. These suggest that the proportions of observations whose informational efficiency increases or decreases are approximately 50/50. Moreover, the mean and median values of $\Delta VR(6)$ and $\Delta VR(12)$ all equal zero, suggesting similar proportions of stocks whose prices have become more efficient or less efficient. The descriptive results of proxies for stock price informational efficiency are quite similar, indicating the reliability of our results.

Additionally, the mean value of changes in share turnover (Δ*Turn*) equals zero, and the minimal and maximum values of turnover changes are −0.06 and 0.05, respectively. The mean percentage of changes in the *Amivest* measure is 0.06, with the level variable being 9.2. For the *Gamma* variable, the mean values of the level and change variables are −0.05 and − 0.01, respectively.

4. Empirical results

4.1. ETF ownership and informational efficiency

4.1.1. Non-synchronicity, dynamic volume–return relation, and variance ratio tests

To test the effect of ETF activity on a firm's stock price informational efficiency, we first estimate the following equation[4]:

$$\begin{aligned} \Delta Info_{it} = {} & \beta_0 + \beta_1 \Delta ETF_{i,t-1} + \beta_2 \Delta Int_{i,t-1} + \beta_3 \Delta Beta_{i,t-1} + \beta_4 \Delta BTM_{i,t-1} \\ & + \beta_5 \Delta Log\left(_m + \beta_6 \Delta Skew_{i,t-1} + \beta_7 \Delta Turn_{i,t-1} + \beta_8 ETFdummy_{i,t-1} \right. \\ & + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_l + \varepsilon_{it} \end{aligned} \tag{14}$$

In Eq. (14), $\Delta info_{it}$ is the difference between firm i's measure of informational efficiency during year t and its value in year t-1. We use five variables to proxy for stock price informational efficiency: *NonSynch*, $LMSW_{res}$, $LMSW_{dif}$, *VR(6),* and *VR(12)*. Detailed definitions of these five variables can be seen in Section 3.1.1. The variable of interest, $\Delta ETF_{i,\,t-1}$, is the change in the percentage of firm i's shares held by all equity-typed ETFs from the end of year t-1 to the end of year t-2. H1a predicts that coefficient β_1 is positive when we use *Nonsynch*, $LMSW_{res}$, and $LMSW_{dif}$ as proxies for informational efficiency, and it is negative when we use *VR(6)* and *VR(12)*. This

[4] We design our empirical tests using annual panels because we expect the effect of increased ETF ownership to manifest itself gradually over time after an increase in ETF ownership.

Table 1
Summary statistics.

Panel A: Descriptive statistics for change variables

Variable	Obs.	Mean	Std	Min	P25	P50	P75	Max
ΔNonsynch	28,724	0.05	0.90	−2.11	−0.51	0.05	0.60	2.37
$\Delta LMSW_{dif}$	28,724	0.00	0.20	−0.44	−0.12	0.00	0.12	0.44
$\Delta LMSW_{res}$	28,724	0.00	0.21	−0.46	−0.12	0.00	0.13	0.47
ΔVR(12)	28,724	0.00	0.56	−1.50	−0.14	0.00	0.14	1.51
ΔVR(6)	28,724	0.00	0.31	−0.93	−0.10	0.00	0.11	0.95
ΔETF (%)	28,724	0.22	2.25	−6.05	−0.12	0.01	0.42	8.11
ΔETF_{szc} (%)	28,724	0.02	0.49	−1.44	0.00	0.00	0.02	2.11
ΔETF_{shc} (%)	28,724	0.23	1.36	−3.01	0.00	0.00	0.06	5.38
ΔETF_{szs} (%)	28,724	−0.04	1.25	−4.24	0.00	0.00	0.00	3.28
ΔETF_{shs} (%)	28,724	0.00	0.75	−2.18	0.00	0.00	0.00	2.23
ΔTurn	28,724	0.00	0.02	−0.06	−0.01	0.00	0.01	0.05
ΔAmivest	28,724	0.06	0.78	−1.65	−0.48	0.06	0.60	1.89
ΔGamma	28,724	−0.01	0.28	−0.67	−0.10	0.00	0.09	0.68
ΔVPIN(8)	28,724	0.00	0.02	−0.04	−0.01	0.00	0.01	0.04
ΔVPIN(50)	28,724	0.00	0.03	−0.07	−0.02	0.00	0.01	0.07
ΔInt (%)	28,724	0.78	255.60	−29.71	−2.92	−0.13	1.65	24.39
ΔBeta	28,724	−0.01	0.49	−1.09	−0.18	0.00	0.18	0.87
ΔBTM	28,724	0.02	0.19	−0.47	−0.09	0.03	0.13	0.50
ΔLog(mve)	28,724	0.18	0.62	−1.15	−0.23	0.10	0.52	1.86
ΔSkew	28,724	−0.26	1.49	−7.57	−0.42	−0.02	0.33	1.70
ΔStdret	28,724	0.00	0.04	−0.07	−0.01	0.00	0.00	0.03
ΔIntan	28,724	0.00	0.03	−0.07	0.00	0.00	0.00	0.10

Panel B: Descriptive statistics for level variables

Variable	Obs	Mean	Std	Min	P25	P50	P75	Max
Nonsynch	28,724	0.39	0.84	−1.49	−0.16	0.36	0.88	2.63
$LMSW_{dif}$	28,724	−0.06	0.15	−0.36	−0.14	−0.06	0.03	0.27
$LMSW_{res}$	28,724	−0.06	0.16	−0.38	−0.15	−0.06	0.03	0.28
VR(12)	28,724	0.26	0.39	0.00	0.09	0.19	0.31	1.75
VR(6)	28,724	0.18	0.22	0.00	0.06	0.13	0.22	1.14
ETF (%)	28,724	1.96	3.06	0.00	0.06	0.48	2.74	13.98
ETF_{szc} (%)	28,724	0.22	0.52	0.00	0.00	0.03	0.17	2.77
ETF_{shc} (%)	28,724	0.94	1.87	0.00	0.00	0.02	1.34	8.23
ETF_{szs} (%)	28,724	0.47	1.56	0.00	0.00	0.00	0.07	7.88
ETF_{shs} (%)	28,724	0.33	1.09	0.00	0.00	0.00	0.05	5.31
Turn	28,724	0.02	0.02	0.00	0.01	0.02	0.03	0.09
Amivest	28,724	9.20	1.07	6.93	8.47	9.14	9.86	12.06
Gamma	28,724	−0.05	0.21	−0.61	−0.12	−0.04	0.03	0.44
VPIN(8)	28,724	0.18	0.02	0.14	0.16	0.18	0.19	0.23
VPIN(50)	28,724	0.31	0.04	0.19	0.29	0.31	0.33	0.37
Ret	28,724	0.16	0.63	−0.71	−0.24	0.01	0.37	2.44
Earn	28,724	0.02	0.16	−0.36	0.00	0.02	0.04	0.19
EarnAgg	28,724	0.07	0.08	−0.12	0.07	0.08	0.12	0.21
EarnFirm	28,724	−0.06	0.17	−0.43	−0.10	−0.06	−0.02	0.21
Int (%)	28,724	9.12	255.65	0.00	0.85	3.81	10.16	49.54
Beta	28,724	1.16	0.28	0.46	0.99	1.16	1.32	1.96
BTM	28,724	0.55	0.27	0.08	0.34	0.52	0.74	1.15
Log(mve)	28,724	15.24	1.15	12.93	14.47	15.14	15.87	18.69
Skew	28,724	0.04	0.51	−1.15	−0.22	0.01	0.29	1.23
ATGrowth	28,724	1.30	5.52	0.62	1.00	1.09	1.21	3.15
Stdret	28,724	0.03	0.01	0.01	0.02	0.03	0.03	0.06
Intan	28,724	0.05	0.07	0.00	0.01	0.03	0.06	0.33
Loss	28,724	0.16	0.37	0.00	0.00	0.00	0.00	1.00
Mom	28,724	0.04	0.54	−1.25	−0.31	0.01	0.37	1.42

This table presents the descriptive statistics for the main variables in our analyses. Panels A and B display results for the variables in change and level forms, respectively. The sample period is from 2006 to 2020. Variable definitions are provided in Appendix Table A.1.

indicates that, ceteris paribus, increases in ETF ownership are associated with increases in stock price informational efficiency. In contrast, H1b predicts that coefficient β_1 is negative when we proxy for informational efficiency using *Nonsynch*, $LMSW_{res}$, and $LMSW_{dif}$, and it is positive when we use the *VR(6)* and *VR(12)* measures. This indicates that, ceteris paribus, increases in ETF ownership are associated with decreases in stock price informational efficiency. Following Israeli et al. (2017), we use change forms of variables to eliminate the stickiness of informational efficiency and ETF ownership, thereby alleviating potential endogeneity problems.

To isolate the effect of changes in ETF ownership on informational efficiency and ensure our results are not confounded by the relation between ETF ownership and institutional ownership, we include $\Delta Int_{i,\,t-1}$ directly in Eq. (14) as an additional variable. $\Delta Int_{i,\,t-1}$ is the change in the percentage of firm i's shares held by all institutions from the end of year t-2 to the end of year t-1. Following prior literature (Jin and Myers, 2006; Li et al., 2014), we also include the change in stock return skewness during year t-1 ($\Delta Skew_{i,\,t-1}$) and the change in the capital asset pricing model (CAPM) beta during year t-1 ($\Delta Beta_{i,\,t-1}$) to control for the asymmetric distribution of stock returns and a firm's systematic risk. $Beta_{it}$ is calculated by the following regression analysis:

$$R_{id} = \alpha + Beta_{it} \times R_d^M + \varepsilon_{id} \tag{15}$$

where R_{id} and R_d^M are stock i's return and value-weighted market return on day d, respectively. $Beta_{it}$ is the coefficient estimate before R_d^M. Eq. (15) is estimated separately for each stock year, using daily returns of stock i over the trading days in year t.

In addition, the control variables also include the change in the log of the market value of equity [ΔLog(mve)] during year t-1 because larger firms generally have higher informational efficiency (Israeli et al., 2017). We also include the change in the book-to-market ratio (Δ*BTM*) during year t-1 as a control for the effect of financial distress and growth opportunity (Fama and French, 1992; Lakonishok et al., 1994), and we include the change in turnover (Δ*Turn*) during year t-1 as a control for the effect of stock liquidity. Considering that the strategy of most ETFs is to follow passive indexes, becoming the underlying component securities of an ETF may directly change the components' information efficiencies. Therefore, we also include the *ETFdummy* variable in year t-1, which equals one if the stock is an underlying security of the ETF in year t-1; otherwise, it equals zero. Finally, to control time and industry trends in stock price informational efficiency, we include year and industry-fixed effects. The industry-fixed effects are defined based on the GICS industry classification.

Table 2 tabulates the regression analysis results of Eq. (14). Columns (1) to (5) present the regression analysis results with the dependent variables as changes in non-synchronicity (Δ*Nonsynch*), changes in the dynamic relation between volume and return based on regression analysis of difference terms or residual terms ($\Delta LMSW_{dif}$ or $\Delta LMSW_{res}$), and changes in VR based on the 6- or 12-period lags [Δ*VR*(6) or Δ*VR*(12)]. Column (1) reveals that changes in non-synchronicity (Δ*Nonsynch*) exhibit a significant positive relation with ΔETF_{t-1}. Economically, the result shows that a one percentage point increase in ETF ownership is associated with a 12.93 percentage point increase in *Nonsynch* over the next year. This finding supports H1a that informational efficiency increases with changes in ETF ownership. Similar results are presented in columns (2) and (3), with significantly positive relations between ETF activities and $\Delta LMSW_{dif}$ as well as ETF activities and $\Delta LMSW_{res}$. For Δ*VR*(6) and Δ*VR*(12), the results in columns (4) and (5) suggest that increases in ETF ownership are associated with increases in the informational efficiency of stock prices.

Moreover, the *ETFdummy* variable is positively related to informational efficiency, denoting the necessity of controlling the status of underlying securities based on whether they are component securities of ETFs. Taken together, the results in Table 2 provide strong evidence in support for H1a that an increase in ETF ownership is accompanied by an increase in the informational efficiency of underlying stocks.

Table 2
ETF activity and informational efficiency of underlying securities.

	(1)	(2)	(3)	(4)	(5)
	ΔNonsynch	$\Delta LMSW_{dif}$	$\Delta LMSW_{res}$	ΔVR(6)	ΔVR(12)
ΔETF_{t-1}	12.93***	0.72***	0.52***	−2.21***	−3.26***
	(8.97)	(3.81)	(4.23)	(−3.95)	(−3.98)
ΔInt_{t-1}	0.06	0.00	0.02	−0.02	−0.04
	(1.20)	(0.10)	(1.11)	(−1.02)	(−1.47)
$\Delta Beta_{t-1}$	0.30***	−0.01***	−0.02***	−0.01	0.01
	(22.62)	(−3.66)	(−5.99)	(−1.21)	(1.59)
ΔBTM_{t-1}	−0.02	−0.02	−0.02	0.05***	0.08***
	(−0.37)	(−1.35)	(−1.36)	(3.02)	(3.32)
$\Delta Log(mve)_{t-1}$	−0.29***	0.00	0.01*	−0.08***	−0.12***
	(−25.46)	(1.34)	(1.95)	(−18.70)	(−19.14)
$\Delta Skew_{t-1}$	0.00	0.00**	0.00	−0.00	−0.00
	(0.23)	(2.37)	(1.52)	(−0.85)	(−0.77)
$\Delta Turn_{t-1}$	−5.94***	−0.06	−0.02	−1.15***	−1.53***
	(−26.90)	(−0.99)	(−0.26)	(−13.39)	(−12.21)
$ETFdummy_{t-1}$	0.07***	0.00***	0.00***	−0.01**	−0.01**
	(8.28)	(2.95)	(3.05)	(−2.34)	(−2.07)
Constant	−0.47***	−0.00	−0.01	0.03**	0.05**
	(−14.56)	(−0.06)	(−1.52)	(2.57)	(2.52)
Observations	28,422	28,390	28,390	28,408	28,408
R-squared	0.463	0.037	0.030	0.070	0.063
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

This table presents regression analysis results of changes in informational efficiency (i.e., *ΔNonsynch*, *ΔLMSW(dif)*, *ΔLMSW(res)*, *ΔVR(6)*, and *ΔVR(12)*) on changes in ETF ownership (*ΔETF*) and other control variables. t-statistics based on standard error and clustered by firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for variable definitions.

4.1.2. Future earnings response coefficient tests

This section explores the influence of ETF activities on the sensitivity of stock returns to fundamental earnings. We follow the literature (Collins et al., 1994; Israeli et al., 2017) and estimate the following regression model:

$$\begin{aligned} Ret_{it} &= \beta_0 + \beta_1 Earn_{i,t-1} + \beta_2 Earn_{i,t} + \beta_3 Earn_{i,t+1} + \beta_4 ETF_{i,t-1} + \beta_5 Ins_{i,t-1} \\ &+\beta_6 ETF_{i,t-1} \times Earn_{i,t-1} + \beta_7 ETF_{i,t-1} \times Earn_{i,t} + \beta_8 ETF_{i,t-1} \times Earn_{i,t+1} \\ &+\sum_k \beta_k Controls_{i,t-1/t/t+1} + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_j + \varepsilon_{it} \end{aligned} \tag{16}$$

In Eq. (16), Ret_{it} represents firm-level stock returns during year t, and $Earn_{i,\,t-1}$, $Earn_{i,\,t}$, and $Earn_{i,\,t+1}$ respectively denote firm-level net income before extraordinary items during year t-1, t, and t + 1, scaled by the market value of equity. Coefficients β_1, β_2, and β_3 measure the relation between current firm-level stock returns and past, current, and future firm earnings. A positive

Table 3
ETF activity and future earning response coefficients.

	(1)	(2)
	Ret_t	Ret_t
$Earn_{t-1}$	0.01	−0.03
	(0.17)	(−0.60)
$Earn_t$	0.49***	0.51***
	(10.00)	(10.13)
$Earn_{t+1}$	0.27***	0.26***
	(5.78)	(5.29)
ETF_{t-1}	0.49	
	(0.59)	
ΔETF_{t-1}		1.73
		(1.57)
ETF_{t-2}		−0.48
		(−0.52)
$ETF_{t-1} \times Earn_{t-1}$	28.79***	
	(2.58)	
$ETF_{t-1} \times Earn_t$	−10.26	
	(−0.06)	
$ETF_{t-1} \times Earn_{t+1}$	21.95***	
	(2.59)	
$\Delta ETF_{t-1} \times Earn_{t-1}$		−5.47
		(−0.36)
$\Delta ETF_{t-1} \times Earn_t$		0.73
		(0.06)
$\Delta ETF_{t-1} \times Earn_{t+1}$		12.86**
		(2.09)
$ETF_{t-2} \times Earn_{t-1}$		59.58***
		(4.10)
$ETF_{t-2} \times Earn_t$		−25.21
		(−0.93)
$ETF_{t-2} \times Earn_{t+1}$		34.04***
		(2.77)
Ins_{t-1}	−0.17***	−0.17***
	(−6.54)	(−6.51)
$Log(mve)_{t-1}$	−0.08***	−0.08***
	(−23.90)	(−23.76)
$Loss_t$	−0.08***	−0.08***
	(−9.14)	(−9.14)
$ATGrowth_t$	0.10***	0.10***
	(19.22)	(19.24)
Ret_{t+1}	−0.10***	−0.10***
	(−15.51)	(−15.50)
Constant	0.68***	0.68***
	(11.82)	(11.89)
Observations	27,082	27,082
R-squared	0.543	0.544
Year FE	YES	YES
Industry FE	YES	YES

This table presents regression analysis results of current annual stock return (Ret_t) on total earnings in the past, current, and future periods ($Earn_{t-1}$, $Earn_t$, and $Earn_{t+1}$); lagged level of ETF ownership (ETF_{t-1}); lagged changes in ETF ownership (ΔETF_{t-1}); and several interactions between total earnings and lagged ETF ownership. t-statistics based on standard error and clustered by firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for variable definitions.

FERC—β_3—reflects the informativeness of the stock price. $Inst_{i,\ t-1}$ is the level of institutional ownership, which is controlled to isolate the effect of institutional ownership on stock returns. To address our main research question, we include as explanatory variables the level of ETF ownership ($ETF_{i,\ t-1}$) at the end of year t-1 as well as the interactions between the level of ETF ownership and past, current, and future earnings ($ETF_{i,\ t-1} \times Earn_{i,\ t\pm j}$). H1a predicts that the FERC is higher for firms with higher ETF ownership (i.e., β_8 is positive). In contrast, H1b predicts that the FERC is lower for firms with higher ETF ownership (i.e., β_8 is negative).

$Control_{i,t\text{-}1/t/t+1}$ denotes several control variables suggested by prior research. Following Collins et al. (1994), we control for future firm-level stock returns, $Ret_{i,\ t+1}$, to address the potential measurement error induced by using actual future earnings as a proxy for expected future earnings. In addition, to account for the effect of a firm's growth on the ability of its stock returns to reflect future earnings, we control for total asset growth during year t, $ATGrowth_{it}$. Additionally, firms experiencing losses may have significantly different FERCs; thus, we include an indicator variable, $Loss_{it}$, that equals 1 if the firm experiences a loss in year t and 0 otherwise. In addition, $Control_{i,t\text{-}1/t/t+1}$ also includes the level of institutional ownership in year t-1 (Int_{t-1}) and the natural logarithm of the market value of equity at the end of year t [Log(mve)]. Similarly, we also include year and industry-fixed effects to control for time and industry trends in stock returns.

We further decompose the level of ETF ownership at the end of year t-1 into the sum of the level of ETF ownership at the end of year t-2 and the changes in ETF ownership during year t-1. By definition, $ETF_{t-1} = ETF_{t-2} + \Delta ETF_{t-1}$. Specifically, we estimate Eq. (17) using ETF_{t-2} and ΔETF_{t-1} instead of ETF_{t-1}:

$$\begin{aligned}Ret_{it} = {} & \beta_0 + \beta_1 Earn_{i,t-1} + \beta_2 Earn_{i,t} + \beta_3 Earn_{i,t+1} + \beta_4 ETF_{i,t-2} + \beta_5 \Delta ETF_{i,t-1} + \beta_6 Inst_{i,t-1} \\ & + \beta_7 ETF_{i,t-2} \times Earn_{i,t-1} + \beta_8 ETF_{i,t-2} \times Earn_{i,t} + \beta_9 ETF_{i,t-2} \times Earn_{i,t+1} \\ & + \beta_{10} \Delta ETF_{i,t-1} \times Earn_{i,t-1} + \beta_{11} \Delta ETF_{i,t-1} \times Earn_{i,t} + \beta_{12} \Delta ETF_{i,t-1} \times Earn_{i,t+1} \\ & + \sum_k \beta_k Controls_{i,t-1/t/t+1} + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_l + \varepsilon_{i,t}\end{aligned} \tag{17}$$

In estimating Eq. (17), we are interested in the coefficients on the interactions of the lagged levels of ETF ownership with future firm earnings (i.e., β_9) as well as the coefficients on the interactions of the lagged changes in ETF ownership with future firm earnings (i.e., β_{12}). H1a (H1b) predicts that the coefficients on these interaction terms are positive (negative).

Table 3 presents the regression analysis results from the estimation of Eqs. (16) and (17). Consistent with the literature, in both columns, we observe positive FRECs (coef. $= 0.027$, $t = 5.78$ and coef. $= 0.026$, $t = 5.29$). More importantly, consistent with H1a, Column (1) reveals that the interaction of future earnings with ETF ownership carries a positive coefficient that is significantly different from zero (coef. $= 21.95$, $t = 2.59$). This suggests that firms with higher levels of ETF ownership experience higher FERCs. Column (2) presents results from the estimation of Eq. (17), in which the level of ETF ownership is split into a lagged level of ETF ownership and the most recent period change in ETF ownership. The results in Column (2) further support H1a by showing that the coefficients on the interactions of future earnings with both the change and lagged level in ETF are also significantly positive (coef. $=$ 12.86, $t = 2.09$ and coef. $= 34.04$, $t = 2.77$).

Taken together, the results presented in Table 3 indicate that firms experiencing increases in ETF ownership also experience increases in the average magnitude of their FERCs and thus increase their pricing efficiency. This supports H1a.

The previous results show that ETF activity results in increases in informational efficiency for underlying stocks. The next question is whether such improvement should be attributable to firm-specific information or macro-based information, or both. To answer this question, first, we follow prior literature (Israeli et al., 2017; Glosten et al., 2021) and decompose total earnings into "macro-based" (*EarnAgg*) and "firm-specific" (*EarnFirm*) components. We estimate these two components using the model in Eq. (18), where *EarnAgg* is calculated as the fitted value from the annual regression analysis for each stock, and *EarnFirm* is obtained as the regression analysis residual.

$$Earn_{i,t} = \beta_0 + \beta_1 EarnMkt_t + \beta_2 EarnInd_{i,t} + \varepsilon_{it} \tag{18}$$

where $EarnMkt_t$ is the value-weighted average of *Earn* of all shares in year t and $EarnInd_{i,\ t}$ is the value-weighted average of *Earn* of all firms in the same industry as firm i in year t.

We then estimate the following equation by regressing annual stock returns on systemic and firm-specific earnings news and regressing their interactions with the level of ETF ownership in year t-1:

$$\begin{aligned}Ret_{it} = {} & \beta_0 + \beta_1 EarnAgg_{i,t-1} + \beta_2 EarnFirm_{i,t-1} + \beta_3 EarnAgg_{i,t} + \beta_4 EarnFirm_{i,t} \\ & + \beta_5 EarnAgg_{i,t+1} + \beta_6 EarnFirm_{i,t+1} + \beta_7 ETF_{i,t-1} + \beta_8 Ins_{i,t-1} \\ & + \beta_9 ETF_{i,t-1} \times EarnAgg_{i,t-1} + \beta_{10} ETF_{i,t-1} \times EarnAgg_{i,t} \\ & + \beta_{11} ETF_{i,t-1} \times EarnAgg_{i,t+1} + \beta_{12} ETF_{i,t-1} \times EarnFirm_{i,t-1} \\ & + \beta_{13} ETF_{i,t-1} \times EarnFirm_{i,t} + \beta_{14} ETF_{i,t-1} \times EarnFirm_{i,t+1} \\ & + \sum_k \beta_k Controls_{i,t-1/t/t+1} + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_j + \varepsilon_{it}\end{aligned} \tag{19}$$

We are interested in the coefficients on the interactions of lagged ETF with $EarnAgg_{i,\ t+1}$ and lagged ETF with $EarnFirm_{i,\ t+1}$ (i.e., β_{11} and β_{14}), which indicates how macro-based and firm-specific FERCs vary with the lagged level of ETF ownership.

We also re-estimate Eq. (19) using ETF_{t-2} and ΔETF_{t-1} in place of ETF_{t-1}. The model specification is similar to Eq. (17) except that we decompose total earnings into "macro-based" (*EarnAgg*) and "firm-specific" (*EarnFirm*) components.

Table 4 suggests that the increase in pricing efficiency is attributable to firm-specific earnings information. Specifically, in Column

Table 4
ETF activity and future earning response coefficients after decomposing total earnings.

	(1)	(2)
	Ret_t	Ret_t
$EarnAgg_{t-1}$	2.32***	2.25***
	(3.94)	(3.81)
$EarnFirm_{t-1}$	−0.02	−0.06
	(−0.33)	(−0.97)
$EarnAgg_t$	−1.52***	−1.41**
	(−2.61)	(−2.42)
$EarnFirm_t$	0.48***	0.50***
	(9.85)	(9.97)
$EarnAgg_{t+1}$	2.12***	2.40***
	(4.30)	(4.85)
$EarnFirm_{t+1}$	0.24***	0.22***
	(5.18)	(4.57)
ETF_{t-1}	0.88	
	(0.54)	
ΔETF_{t-1}		−1.15
		(−0.44)
ETF_{t-2}		4.04**
		(2.20)
$ETF_{t-1} \times EarnAgg_{t-1}$	78.12***	
	(5.17)	
$ETF_{t-1} \times EarnFirm_{t-1}$	23.63**	
	(2.06)	
$ETF_{t-1} \times EarnAgg_t$	−35.16	
	(−0.19)	
$ETF_{t-1} \times EarnFirm_t$	4.33	
	(0.44)	
$ETF_{t-1} \times EarnAgg_{t+1}$	24.78	
	(0.41)	
$ETF_{t-1} \times EarnFirm_{t+1}$	30.13***	
	(3.54)	
$\Delta ETF_{t-1} \times EarnAgg_{t-1}$		0.12
		(0.00)
$\Delta ETF_{t-1} \times EarnFirm_{t-1}$		−5.40
		(−0.36)
$\Delta ETF_{t-1} \times EarnAgg_t$		20.75
		(0.81)
$\Delta ETF_{t-1} \times EarnFirm_t$		12.59
		(0.99)
$\Delta ETF_{t-1} \times EarnAgg_{t+1}$		−20.71
		(−1.38)
$\Delta ETF_{t-1} \times EarnFirm_{t+1}$		15.09**
		(2.28)
$ETF_{t-2} \times EarnAgg_{t-1}$		119.41***
		(6.70)
$ETF_{t-2} \times EarnFirm_{t-1}$		55.98***
		(3.66)
$ETF_{t-2} \times EarnAgg_t$		52.60
		(0.62)
$ETF_{t-2} \times EarnFirm_t$		12.89***
		(2.98)
$ETF_{t-2} \times EarnAgg_{t+1}$		56.72
		(0.38)
$ETF_{t-2} \times EarnFirm_{t+1}$		47.08***
		(3.83)
Ins_{t-1}	−0.16***	−0.16***
	(−6.19)	(−6.08)
$Log(mve)_{t-1}$	−0.08***	−0.08***
	(−23.23)	(−23.31)
$Loss_t$	−0.08***	−0.08***
	(−9.39)	(−9.42)
$ATGrowth_t$	0.10***	0.10***
	(19.24)	(19.30)
Ret_{t+1}	−0.10***	−0.10***
	(−15.24)	(−15.30)
Constant	0.88***	0.89***
	(11.19)	(11.23)
Observations	27,082	27,082

(*continued on next page*)

Table 4 (*continued*)

	(1)	(2)
	Ret_t	Ret_t
R-squared	0.545	0.546
Year FE	FE	FE
Industry FE	YES	YES

This table presents the results of regression analyses of current annual stock return (Ret_t) on aggregate and firm-specific earnings in the past, current, and future periods ($EarnAgg_{t-1}$, $EarnAgg_t$, and $EarnAgg_{t+1}$ as well as $EarnFirm_{-1}$, $EarnFirm_t$, and $EarnFirm_{t+1}$); lagged level of ETF ownership (ETF_{t-1}); lagged changes in ETF ownership (ΔETF_{t-1}); and several interactions between aggregate and firm-specific earnings and lagged ETF ownership. t-statistics based on standard and error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for variable definitions.

(1), the coefficient on the interaction of $ETF_{i,\ t-1} \times EarnFirm_{i,\ t+1}$ is positive and significant (coef. = 30.13; $t = 3.54$), implying that an increase in ETF activity pushes prices to reflect more firm-specific information. The coefficient on the interaction term of $ETF_{i,\ t-1} \times EarnAgg_{i,\ t+1}$ is insignificant, although positive, indicating that ETF activity does not increase pricing information related to systematic information. The results in Column (2) paint a similar picture after we split the level of ETF ownership into a lagged level of ETF ownership and the most recent period change in ETF ownership. The significantly positive coefficients on the interaction terms of $\Delta ETF_{t-1} \times EarnFirm_{t+1}$ (coef. = 15.09; $t = 2.28$) and $ETF_{t-2} \times EarnFirm_{t+1}$ (coef. =47.08; $t = 3.83$) indicate that the contribution of ETF activity to pricing efficiency can be attributed to its pricing discovery in firm-specific information.

Taken together, the results presented in Tables 2, 3, and 4 indicate that an increase in ETF ownership is associated with an increase in the informational efficiency of the underlying stocks. ETF activities not only increase the non-synchronicity of stocks, strengthen the dynamic relation between return and volume, and decrease the VR, but they also cause more future earnings (especially firm-specific earnings news) to be incorporated into stock prices. These findings all support H1a.

4.2. *Endogeneity*

Following Ben David et al. (2018) and Glosten et al. (2021), we employ CSI 300 and CSI 500 index reconstitution as an exogenous shock to ETF ownership and exploit the instrumental variable (IV) method to mitigate the endogeneity problem and corroborate our main findings.

CSI 300/500 index reconstitution offers a setting in which firms with similar characteristics have a significant variation in ETF

Table 5
ETF activity and stock price informational efficiency: IV method.

	(1)	(2)	(3)	(4)	(5)	(6)
	1st stage	2nd stage				
	ΔETF	ΔNonsynch	$\Delta LMSW_{dif}$	$\Delta LMSW_{res}$	ΔVR(6)	ΔVR(12)
ΔETF_{t-1}		14.79**	2.62***	2.23***	−2.42***	−4.53***
		(2.40)	(2.85)	(2.68)	(−2.75)	(−2.94)
$Reconst_t$	0.01***					
	(12.18)					
ΔInt_{t-1}	0.00*	0.20	0.03	0.03	0.01	0.00
	(1.65)	(0.56)	(0.37)	(0.27)	(0.12)	(0.02)
$\Delta Beta_{t-1}$	0.00	0.14	−0.04	−0.03	0.03	−0.01
	(0.97)	(1.33)	(−1.16)	(−0.77)	(0.80)	(−0.18)
ΔBTM_{t-1}	0.00	0.79**	0.18	0.25**	−0.19*	−0.26
	(0.10)	(2.45)	(1.61)	(2.27)	(−1.87)	(−1.61)
$\Delta Log(mve)_{t-1}$	0.00	0.33***	0.01	0.05	−0.04	−0.05
	(0.38)	(3.13)	(0.37)	(1.31)	(−0.99)	(−0.89)
$\Delta Skew_{t-1}$	0.00	0.04	−0.02*	−0.02***	0.02**	0.01
	(1.20)	(1.54)	(−1.93)	(−2.64)	(2.37)	(0.67)
$\Delta Turn_{t-1}$	0.00	−3.05	0.43	0.60	−0.77	−0.18
	(0.11)	(−1.18)	(0.63)	(0.77)	(−0.94)	(−0.15)
$ETFdummy_{t-1}$	−0.00**	−0.04	0.02	0.01	−0.04	−0.01
	(−2.57)	(−0.32)	(0.75)	(0.23)	(−0.97)	(−0.26)
Constant	0.00	−0.43*	−0.00	0.02	0.00	0.19
	(0.97)	(−1.84)	(−0.01)	(0.23)	(0.05)	(1.22)
Observations	482	481	479	479	479	479
R-squared	0.555	0.550	0.234	0.228	0.195	0.178
Year FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES

This table presents the results of the IV method. $Reconst_{i,t}$ is the instrumental variable, which equals 1 if firm i moves from the CSI 300 to the CSI 500 in year t and equals 0 if firm i moves from the CSI 500 to the CSI 300 in year t. t-statistics based on standard error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for variable definitions.

ownership, while it has no direct effect on stock pricing efficiency. Specifically, a market capitalization rank of 300 is the cutoff rank for the CSI 300 index; the next 500 ranked stocks (i.e., the 301st to 800th stocks) constitute the CSI 500 index. Every six months, CSI Index Co., Ltd. makes periodic adjustments to the constituents of the index based on the average daily total market value of the stocks for the most recent year. As the indexes are value weighted, if one of the lowest-market-capitalization stocks of the CSI 300 index switches to the CSI 500 index, ETF ownership of that stock increases significantly. In contrast, if one of the highest-market-capitalization stocks of the CSI 500 index switches to the CSI 300 index, ETF ownership of that stock decreases significantly. Therefore, CSI 300/500 index reconstitution can be used as an IV to test our conjecture that ETF activity increases stock price informational efficiency.

We conduct the following first-stage regression to verify that firms transitioning from the CSI 300 to the CSI 500 experience an increase in ETF ownership relative to firms transitioning from the CSI 500 to the CSI 300.

$$\begin{aligned}\Delta ETF_{it} = \beta_0 + \beta_1 Reconst_{i,t} + \beta_2 \Delta Int_{i,t-1} + \beta_3 \Delta Beta_{i,t-1} + \beta_4 \Delta BTM_{i,t-1} \\ + \beta_5 \Delta Log\left(_m + \beta_6 \Delta Skew_{i,t-1} + \beta_7 \Delta Turn_{i,t-1} + \beta_8 ETFdummy_{i,t-1} \right. \\ + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_l + \varepsilon_{it}\end{aligned} \tag{20}$$

where $Reconst_{i,t}$ equals 1 if firm i moves from the CSI 300 to the CSI 500 in year t and equals 0 if firm i moves from the CSI 500 to the CSI 300 in year t.

In the second-stage regression, we use the predicted value of changes in ETF ownership as the key dependent variable and adopt the following model specification.

$$\begin{aligned}\Delta Info_{it} = \beta_0 + \beta_1 \Delta \widehat{ETF}_{i,t-1} + \beta_2 \Delta Int_{i,t-1} + \beta_3 \Delta Beta_{i,t-1} + \beta_4 \Delta BTM_{i,t-1} \\ + \beta_5 \Delta Log\left(_m + \beta_6 \Delta Skew_{i,t-1} + \beta_7 \Delta Turn_{i,t-1} + \beta_8 ETFdummy_{i,t-1} \right. \\ + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_l + \varepsilon_{it}\end{aligned} \tag{21}$$

Table 5 in our current manuscript reports the results. The coefficient of $Reconst_{i,t}$ in column (1) is significantly positive at the 1% level, which suggests that changes in ETF ownership of a stock increase significantly if it switches from the CSI 300 index to the CSI 500 index. Columns (2)–(6) indicate that stocks moved into the CSI 500 index from the CSI 300 index have higher informational efficiency than stocks moved into the CSI 300 index from the CSI 500 index. Given that the ETF ownership of a stock increases significantly if it switches from the CSI 300 index to the CSI 500 index, the above results are consistent with our prior findings.

Table 6
ETF activity and stock liquidity.

	(1)	(2)	(3)
	ΔAmivest	ΔTurn	ΔGamma
ΔETF_{t-1}	4.82***	0.24***	−0.21***
	(4.12)	(7.61)	(−3.50)
ΔInt_{t-1}	−0.40***	−0.01***	0.05***
	(−10.67)	(−6.09)	(4.13)
ΔBTM_{t-1}	−0.37***	−0.00***	0.01
	(−11.22)	(−3.17)	(1.04)
$\Delta Log(mve)_{t-1}$	0.32***	−0.01***	−0.01*
	(35.79)	(−34.36)	(−1.83)
$\Delta Stdret_{t-1}$	3.99***	0.08***	0.22**
	(13.21)	(9.18)	(2.08)
$\Delta Amivest_{t-1}$	−0.29***		
	(−46.67)		
$\Delta Turn_{t-1}$		−0.10***	
		(−18.35)	
$\Delta Gamma_{t-1}$			−0.49***
			(−87.80)
$ETFdummy_{t-1}$	0.11***	0.00***	−0.01**
	(16.01)	(9.64)	(−2.50)
Constant	−1.05***	−0.01***	−0.01
	(−39.92)	(−19.33)	(−1.10)
Observations	28,408	28,408	28,361
R-squared	0.549	0.395	0.228
Year FE	YES	YES	YES
Industry FE	YES	YES	YES

This table presents the results of the regression analysis of changes in stock liquidity (i.e., *ΔAmivest*, *ΔTurn*, and *ΔGamma*) on changes in ETF ownership (*ΔETF*) and other control variables. t-statistics based on standard error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for variable definitions.

4.3. Channel tests

4.3.1. ETF ownership and stock liquidity

In this section, we further explore the underlying channels through which ETF activities increase the informational efficiency of underlying stocks. According to previous studies (Fremault, 1991; Kumar and Seppi, 1994; Holden, 1995; Subrahmanyam and Titman, 1999), ETFs can improve the stock liquidity of underlying stocks through arbitraging, which may further contribute to the higher pricing efficiency of underlying stocks. To investigate the relation between ETF ownership and stock liquidity, we estimate the following regression analysis:

$$\begin{aligned}\Delta Liquidity_{it} = {} & \beta_0 + \beta_1 \Delta ETF_{i,t-1} + \beta_2 \Delta Int_{i,t-1} + \beta_3 \Delta BTM_{i,t-1} + \beta_4 \Delta Log\big(_m \\ & + \beta_5 \Delta Stdret_{i,t-1} + \beta_6 \Delta Liquidity_{i,t-1} + \beta_7 ETFdummy_{i,t-1} \\ & + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy_l + \varepsilon_{it}\end{aligned} \tag{22}$$

where $\Delta Liquidity_{it}$ is the difference between firm i's measure of liquidity in year t and its value in year t-1. As denoted in Section 3.1.3, we use three proxies to measure stock liquidity: (1) share turnover (*Turn*), (2) the Amivest measure (*Amivest*), and (3) the Gamma measure (*Gamma*). The first two are positive indicators; that is, higher values of measures are associated with higher stock liquidity. The last is a negative indicator; that is, the greater the *Gamma*, the worse the stock liquidity.

Changes in ETF ownership may be correlated with changes in institutional ownership. Moreover, prior research suggests that there might be a relation between institutional ownership and liquidity (Glosten and Harris, 1988); thus, we include $\Delta Int_{i,\ t-1}$ directly in the equation as an additional control variable. In addition, we include the change in the book-to-market ratio ($\Delta BTM_{i,\ t-1}$) during year t-1 as a control for the effect of financial distress or growth opportunities on stock liquidity (Lakonishok et al., 1994). We include $\Delta Log(mve)_{i,\ t-1}$, as larger stocks generally have higher liquidity. Prior studies also find that stock liquidity increases with return volatility; thus, we control for the change in the annualized standard deviation of daily returns during year t-1 ($\Delta Stdret_{i,\ t-1}$). To rule out any autocorrelation between stock liquidity, we also include the lag-term of changes in stock liquidity ($\Delta Liquidity_{i,\ t-1}$). Considering the special feature of becoming the underlying component securities of an ETF, we also include the lagged *ETFdummy*. Finally, year and industry-fixed effects are also included.

Table 7
Stock liquidity and informed trading.

	(1)	(2)	(3)	(4)	(5)	(6)
	ΔVPIN(8)			ΔVPIN(50)		
$\Delta Amivest_t$	0.001**			0.001***		
	(2.08)			(3.13)		
$\Delta Turn_t$		0.005***			0.099***	
		(2.86)			(11.40)	
$\Delta Gamma_t$			−0.001***			−0.001***
			(−2.91)			(−2.84)
ΔInt_{t-1}	−0.014***	−0.014***	−0.014***	0.011***	0.010***	0.011***
	(−13.14)	(−13.22)	(−13.27)	(6.14)	(5.58)	(6.40)
Int_{t-2}	−0.010***	−0.010***	−0.010***	0.002*	0.002	0.003*
	(−12.06)	(−12.18)	(−12.23)	(1.69)	(1.23)	(1.94)
$Log(mve)_{t-1}$	−0.000***	−0.000***	−0.000***	−0.001***	−0.001***	−0.001***
	(−2.79)	(−2.68)	(−2.66)	(−5.99)	(−5.91)	(−6.21)
ΔBTM_{t-1}	−0.003***	−0.003***	−0.003***	0.016***	0.015***	0.015***
	(−3.58)	(−3.45)	(−3.45)	(11.36)	(11.09)	(11.19)
$\Delta Turn_{t-1}$	0.010**	0.010**	0.010**	−0.018**	−0.013*	−0.019**
	(2.08)	(2.14)	(2.21)	(−2.28)	(−1.66)	(−2.46)
$\Delta Stdret_{t-1}$	0.014*	0.014*	0.014*	−0.038***	−0.046***	−0.036***
	(1.70)	(1.72)	(1.73)	(−2.73)	(−3.37)	(−2.64)
$\Delta Intan_{t-1}$	−0.001	−0.000	−0.001	−0.001	−0.001	−0.002
	(−0.14)	(−0.10)	(−0.14)	(−0.18)	(−0.17)	(−0.25)
mom_{t-1}	−0.000	−0.000	−0.000	−0.003***	−0.003***	−0.003***
	(−0.24)	(−0.53)	(−0.47)	(−6.22)	(−4.62)	(−5.99)
ETF $dummy_t$	0.001***	0.001***	0.001***	−0.004***	−0.005***	−0.004***
	(4.53)	(4.40)	(4.37)	(−13.13)	(−13.36)	(−12.90)
Constant	−0.002	−0.002	−0.002	0.034***	0.035***	0.034***
	(−0.61)	(−0.59)	(−0.58)	(5.56)	(5.68)	(5.51)
Observations	28,385	28,385	28,384	28,385	28,385	28,384
R-squared	0.363	0.363	0.363	0.222	0.225	0.222
Year FE	YES	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES	YES

This table presents the results of the regression analysis of changes in informed trading (i.e., *ΔVPIN(8)* and *ΔVPIN(50)* on changes in stock liquidity (i.e., *ΔAmivest*, *ΔTurn*, and *ΔGamma*) and other control variables. t-statistics based on standard error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix for variable definitions.

Table 8
Stock liquidity, informed trading, and informational efficiency.

Panel A: ΔNonsynch

	(1)	(2)	(3)	(4)	(5)
$\Delta Amivest_t$	0.34***				
	(49.97)				
$\Delta Turn_t$		10.87***			
		(40.93)			
$\Delta Gamma_t$			−0.16***		
			(−8.78)		
$\Delta VPIN(8)_t$				4.66***	
				(14.18)	
$\Delta VPIN(50)_t$					2.18***
					(10.95)
ΔInt_{t-1}	−0.13***	−0.01	0.05	−0.09*	−0.09*
	(−2.95)	(−0.29)	(1.06)	(−1.80)	(−1.82)
$\Delta Beta_{t-1}$	0.32***	0.33***	0.30***	0.13***	0.13***
	(25.28)	(25.55)	(22.43)	(8.95)	(9.25)
ΔBTM_{t-1}	0.08**	0.05	0.02	0.09**	0.10**
	(2.01)	(1.13)	(0.47)	(2.00)	(2.08)
$\Delta Log(mve)_{t-1}$	−0.36***	−0.19***	−0.28***	−0.08***	−0.10***
	(−33.08)	(−17.09)	(−25.08)	(−4.65)	(−5.69)
$\Delta Skew_{t-1}$	−0.00	−0.01**	0.00	0.00	0.00
	(−0.98)	(−2.05)	(0.65)	(0.83)	(0.88)
$\Delta Turn_{t-1}$	−5.58***	−5.05***	−5.92***	0.24	0.25
	(−26.37)	(−23.44)	(−26.88)	(1.02)	(1.05)
$ETFdummy_{t-1}$	−0.11***	−0.10***	−0.08***	−0.09***	−0.10***
	(−13.04)	(−11.43)	(−8.93)	(−9.20)	(−10.32)
Constant	−0.03	−0.33***	−0.46***	0.44*	0.45*
	(−0.99)	(−10.32)	(−14.29)	(1.81)	(1.82)
Observations	28,379	28,379	28,378	28,193	28,193
R-squared	0.506	0.493	0.464	0.410	0.408
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel B: $\Delta LMSW_{dif}$

	(1)	(2)	(3)	(4)	(5)
$\Delta Amivest_t$	0.01***				
	(4.46)				
$\Delta Turn_t$		0.01***			
		(3.08)			
$\Delta Gamma_t$			−0.18***		
			(−37.53)		
$\Delta VPIN(8)_t$				0.18**	
				(2.17)	
$\Delta VPIN(50)_t$					0.30***
					(5.77)
ΔInt_{t-1}	−0.00	0.00	−0.01	0.01	0.01
	(−0.30)	(0.05)	(−0.63)	(0.81)	(0.43)
$\Delta Beta_{t-1}$	−0.01***	−0.01***	−0.02***	0.01***	0.01***
	(−3.56)	(−3.76)	(−4.89)	(3.23)	(3.56)
ΔBTM_{t-1}	−0.01	−0.01	−0.01	−0.00	−0.01
	(−1.09)	(−1.23)	(−1.32)	(−0.21)	(−0.53)
$\Delta Log(mve)_{t-1}$	0.00	0.00	0.01**	−0.01**	−0.01*
	(0.78)	(1.41)	(2.17)	(−2.13)	(−1.70)
$\Delta Skew_{t-1}$	0.00**	0.00**	0.00**	−0.00	−0.00
	(2.30)	(2.44)	(2.22)	(−1.42)	(−1.47)
$\Delta Turn_{t-1}$	−0.05	−0.06	−0.08	0.03	0.04
	(−0.84)	(−1.00)	(−1.29)	(0.50)	(0.67)
$ETFdummy_{t-1}$	−0.00	−0.00	−0.00	−0.00	−0.00
	(−1.42)	(−1.09)	(−1.51)	(−1.13)	(−0.52)
Constant	0.01	0.00	0.00	0.07	0.06
	(1.20)	(0.00)	(0.28)	(1.44)	(1.30)
Observations	28,390	28,390	28,390	28,279	28,279
R-squared	0.038	0.037	0.083	0.037	0.038
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel C: ΔLMSW$_{res}$					
	(1)	(2)	(3)	(4)	(5)
ΔAmivest$_t$	0.01***				
	(4.04)				
ΔTurn$_t$		0.09***			
		(3.15)			
ΔGamma$_t$			−0.18***		
			(−33.98)		
ΔVPIN(8)$_t$				0.14**	
				(2.53)	
ΔVPIN(50)$_t$					0.59***
					(10.76)
ΔInt$_{t-1}$	0.01	0.01	0.01	0.01	0.00
	(0.75)	(1.03)	(0.48)	(0.71)	(0.10)
ΔBeta$_{t-1}$	−0.02***	−0.02***	−0.03***	0.01***	0.01***
	(−5.89)	(−5.98)	(−7.13)	(2.72)	(3.22)
ΔBTM$_{t-1}$	−0.01	−0.02	−0.02	0.01	0.00
	(−1.16)	(−1.27)	(−1.37)	(0.67)	(0.14)
ΔLog(*mve*)$_{t-1}$	0.00	0.01**	0.01***	−0.01***	−0.01*
	(1.41)	(2.22)	(2.69)	(−2.60)	(−1.76)
ΔSkew$_{t-1}$	0.00	0.00	0.00	−0.00	−0.00
	(1.44)	(1.49)	(1.34)	(−1.49)	(−1.59)
ΔTurn$_{t-1}$	−0.01	−0.01	−0.03	−0.00	0.01
	(−0.12)	(−0.14)	(−0.52)	(−0.07)	(0.19)
ETFdummy$_{t-1}$	−0.00	−0.00	−0.00	−0.00	−0.00
	(−0.44)	(−0.21)	(−0.50)	(−1.56)	(−0.41)
Constant	−0.00	−0.01	−0.01	0.08	0.07
	(−0.35)	(−1.35)	(−1.26)	(1.64)	(1.43)
Observations	28,390	28,390	28,390	28,279	28,279
R-squared	0.031	0.030	0.068	0.027	0.031
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel D: ΔVR(6)					
	(1)	(2)	(3)	(4)	(5)
ΔAmivest$_t$	−0.07***				
	(−26.34)				
ΔTurn$_t$		−2.93***			
		(−27.97)			
ΔGamma$_t$			0.11***		
			(15.58)		
ΔVPIN(8)$_t$				−2.04***	
				(−16.21)	
ΔVPIN(50)$_t$					−2.18***
					(−28.88)
ΔInt$_{t-1}$	−0.06***	−0.04**	−0.02	−0.06***	−0.02
	(−3.20)	(−1.99)	(−0.87)	(−3.00)	(−0.90)
ΔBeta$_{t-1}$	−0.00	0.00	−0.01	0.01**	0.00
	(−0.29)	(0.49)	(−1.00)	(2.29)	(0.39)
ΔBTM$_{t-1}$	0.07***	0.06***	0.05***	−0.00	0.03
	(4.14)	(3.78)	(3.32)	(−0.23)	(1.55)
ΔLog(*mve*)$_{t-1}$	−0.10***	−0.06***	−0.08***	0.00	−0.01*
	(−22.40)	(−12.89)	(−18.84)	(0.34)	(−1.70)
ΔSkew$_{t-1}$	−0.00	−0.00**	−0.00	0.00	0.00
	(−1.56)	(−2.53)	(−0.60)	(0.13)	(0.41)
ΔTurn$_{t-1}$	−1.07***	−0.91***	−1.14***	−0.04	−0.12
	(−12.66)	(−10.71)	(−13.35)	(−0.41)	(−1.33)
ETFdummy$_{t-1}$	−0.02***	−0.01***	−0.01**	−0.01***	−0.02***
	(−4.61)	(−4.18)	(−2.49)	(−3.06)	(−6.10)
Constant	0.13***	0.07***	0.03***	0.01	0.06
	(9.67)	(5.64)	(2.62)	(0.16)	(0.92)
Observations	28,408	28,408	28,407	28,285	28,285
R-squared	0.091	0.094	0.077	0.045	0.064
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

(*continued on next page*)

Table 8 (*continued*)

Panel E: ΔVR(12)

	(1)	(2)	(3)	(4)	(5)

Panel E: ΔVR(12)

	(1)	(2)	(3)	(4)	(5)
ΔAmivest$_t$	−0.10***				
	(−25.77)				
ΔTurn$_t$		−3.65***			
		(−23.74)			
ΔGamma$_t$			0.12***		
			(11.49)		
ΔVPIN(8)$_t$				−2.61***	
				(−14.31)	
ΔVPIN(50)$_t$					−2.83***
					(−25.79)
ΔInt$_{t-1}$	−0.09***	−0.06**	−0.04	−0.08***	−0.03
	(−3.61)	(−2.31)	(−1.39)	(−3.07)	(−1.19)
ΔBeta$_{t-1}$	0.02**	0.02***	0.01*	0.00	−0.01
	(2.51)	(3.02)	(1.69)	(0.31)	(−1.41)
ΔBTM$_{t-1}$	0.10***	0.09***	0.08***	0.01	0.05*
	(4.42)	(4.00)	(3.61)	(0.34)	(1.94)
ΔLog(*mve*)$_{t-1}$	−0.14***	−0.09***	−0.12***	0.01	−0.01
	(−22.75)	(−14.08)	(−19.17)	(0.86)	(−0.96)
ΔSkew$_{t-1}$	−0.00	−0.00**	−0.00	−0.00	−0.00
	(−1.45)	(−2.16)	(−0.54)	(−0.98)	(−0.74)
ΔTurn$_{t-1}$	−1.42***	−1.24***	−1.52***	0.14	0.04
	(−11.48)	(−9.90)	(−12.17)	(1.08)	(0.28)
ETFdummy$_{t-1}$	−0.02***	−0.02***	−0.01**	−0.02***	−0.03***
	(−4.30)	(−3.67)	(−2.27)	(−3.51)	(−6.23)
Constant	0.18***	0.09***	0.05***	0.02	0.09
	(9.48)	(5.15)	(2.60)	(0.21)	(0.89)
Observations	28,408	28,408	28,407	28,285	28,285
R-squared	0.084	0.080	0.067	0.036	0.051
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

This table presents the results of the regression analysis of changes in informational efficiency (i.e., *ΔNonsynch*, $\Delta LMSW_{dif}$, $\Delta LMSW_{res}$, *ΔVR(6)*, and *ΔVR(12)*) on changes in stock liquidity (i.e., *ΔAmivest*, *ΔTurn*, and *ΔGamma*) and changes in informed trading (i.e., *ΔVPIN(8)* and *ΔVPIN(50)*). Panels A–E are the corresponding results when the proxy for informational efficiency is changed to *ΔNonsynch*, $\Delta LMSW_{dif}$, $\Delta LMSW_{res}$, *ΔVR(6)*, and *ΔVR(12)*, respectively. t-statistics based on standard error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix for variable definitions.

Table 6 presents regression analysis results from the estimation of Eq. (22). Column (1) reveals that the changes in the *Amivest* measure exhibit the expected positive relation with $\Delta ETF_{i,\ t-1}$. This finding indicates that stock liquidity increases with ETF activity. Similarly, in Column (2), the coefficient on $\Delta ETF_{i,\ t-1}$ is significantly positive (coef. = 0.24; $t = 7.61$), indicating the positive effect of ETF ownership on stock liquidity. Column (3) shows that an increase in ETF ownership is associated with a decrease in the *Gamma* measure and thus an increase in stock liquidity.

All results in Table 6 paint the same picture, further strengthening our prior finding that an increase in ETF ownership is accompanied by an increase in stock liquidity, which may facilitate the informational efficiency of underlying securities.

4.3.2. Stock liquidity and informed trading

As suggested by Admati and Pfleiderer (1988) and Foster and Viswanathan (1990), liquidity trading in the market can induce other traders to produce and trade on private information, thereby facilitating the price discovery of securities. To test the relation between stock liquidity and informed investors, we estimate the following regression analysis:

$$\begin{aligned} \Delta Informed_{it} = {} & \beta_0 + \beta_1 \Delta Liquidity_{i,t} + \beta_2 \Delta Int_{it-1} + \beta_3 Int_{it-2} + \beta_4 Log\big(_{m} \\ & + \beta_5 \Delta BTM_{i,t-1} + \beta_6 \Delta Turn_{i,t-1} + \beta_7 \Delta Stdret_{i,t-1} + \beta_8 \Delta Intan_{i,t-1} \\ & + \beta_9 Mom_{i,t-1} + \beta_{10} ETFdummy_{i,t-1} + \sum_j \beta_j Inddummy_j \\ & + \sum_l \beta_l Yeardummy_l + \varepsilon_{it} \end{aligned} \tag{23}$$

where $\Delta Informed_{it}$ is the changes in informed trading, proxied by *VPIN(8)* and *VPIN(50)*.

Table 7 presents the regression analysis results of changes in informed trading on changes in stock liquidity. Columns (1) to (3) show the results when using changes in *VPIN(8)* as the dependent variable, and columns (4) to (6) are the corresponding results of changes in *VPIN(50)*. The significantly positive estimates of the coefficients of $\Delta Amivest_{i,\ t-1}$ and $\Delta Turn_{i,\ t-1}$ as well as the significantly

Table 9
Different types of ETF activity and informational efficiency.

Panel A: ΔNonsynch

	(1)	(2)	(3)	(4)	(5)
$\Delta ETF_{SZCt\text{-}1}$	−25.86				8.50
	(−0.09)				(0.83)
$\Delta ETF_{SHCt\text{-}1}$		16.73***			10.79***
		(5.45)			(3.78)
$\Delta ETF_{SZSt\text{-}1}$			16.31***		1.11**
			(5.57)		(2.31)
$\Delta ETF_{SHSt\text{-}1}$				31.42***	3.73***
				(5.58)	(2.58)
$\Delta Int_{t\text{-}1}$	0.05	0.05	0.05	0.04	0.45***
	(0.98)	(1.04)	(0.98)	(0.95)	(10.12)
$\Delta Beta_{t\text{-}1}$	0.30***	0.30***	0.30***	0.30***	0.28***
	(22.22)	(22.49)	(22.20)	(22.30)	(21.00)
$\Delta BTM_{t\text{-}1}$	0.00	0.00	0.00	−0.00	−0.21***
	(0.06)	(0.05)	(0.03)	(−0.02)	(−4.93)
$\Delta Log(mve)_{t\text{-}1}$	−0.28***	−0.28***	−0.28***	−0.28***	−0.45***
	(−24.90)	(−25.20)	(−24.93)	(−25.16)	(−29.52)
$\Delta Skew_{t\text{-}1}$	0.00	0.00	0.00	0.00	0.01**
	(0.52)	(0.44)	(0.43)	(0.43)	(2.49)
$\Delta Turn_{t\text{-}1}$	−5.95***	−5.95***	−5.94***	−5.93***	−4.36***
	(−26.91)	(−26.94)	(−26.87)	(−26.84)	(−20.17)
$ETFdummy_{t\text{-}1}$	0.08***	0.07***	0.08***	0.08***	−0.07***
	(8.80)	(8.15)	(9.20)	(8.77)	(−7.74)
Constant	−0.46***	−0.46***	−0.47***	−0.46***	−0.61**
	(−14.28)	(−14.21)	(−14.54)	(−14.17)	(−2.57)
Observations	28,422	28,422	28,422	28,422	28,422
R-squared	0.461	0.462	0.462	0.462	0.466
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel B: ΔLMSW(dif)

	(1)	(2)	(3)	(4)	(5)
$\Delta ETF_{SZCt\text{-}1}$	−3.49				1.98
	(−0.52)				(0.69)
$\Delta ETF_{SHCt\text{-}1}$		0.78***			1.26***
		(2.93)			(3.57)
$\Delta ETF_{SZSt\text{-}1}$			0.44***		1.60***
			(3.55)		(3.60)
$\Delta ETF_{SHSt\text{-}1}$				3.04**	0.05**
				(1.97)	(2.03)
$\Delta Int_{t\text{-}1}$	0.00	0.00	0.00	0.00	0.00
	(0.08)	(0.07)	(0.05)	(0.06)	(0.20)
$\Delta Beta_{t\text{-}1}$	−0.01***	−0.01***	−0.01***	−0.01***	−0.01***
	(−3.71)	(−3.68)	(−3.75)	(−3.70)	(−3.35)
$\Delta BTM_{t\text{-}1}$	−0.01	−0.01	−0.01	−0.01	0.02
	(−1.31)	(−1.26)	(−1.25)	(−1.31)	(1.52)
$\Delta Log(mve)_{t\text{-}1}$	0.00	0.00	0.00	0.00	0.02***
	(1.46)	(1.40)	(1.45)	(1.36)	(4.89)
$\Delta Skew_{t\text{-}1}$	0.00**	0.00**	0.00**	0.00**	0.00**
	(2.43)	(2.42)	(2.43)	(2.39)	(2.20)
$\Delta Turn_{t\text{-}1}$	−0.06	−0.06	−0.06	−0.06	−0.09
	(−1.00)	(−1.00)	(−0.99)	(−0.97)	(−1.48)
$ETFdummy_{t\text{-}1}$	−0.00	−0.00	−0.00	−0.00	−0.00
	(−1.01)	(−0.95)	(−1.11)	(−1.02)	(−1.38)
Constant	−0.00	0.00	−0.00	0.00	−0.00
	(−0.03)	(0.01)	(−0.03)	(0.03)	(−0.09)
Observations	28,390	28,390	28,390	28,390	28.390
R-squared	0.037	0.037	0.037	0.037	0.039
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel C: ΔLMSW(res)

(*continued on next page*)

Table 9 (*continued*)

Panel C: ΔLMSW(res)

	(1)	(2)	(3)	(4)	(5)
	(1)	(2)	(3)	(4)	(5)
ΔETF_{SZCt-1}	−4.15				3.37
	(−0.70)				(1.11)
ΔETF_{SHCt-1}		0.90***			1.39***
		(3.00)			(3.63)
ΔETF_{SZSt-1}			0.59***		1.54***
			(3.69)		(3.46)
ΔETF_{SHSt-1}				0.64***	0.71**
				(3.39)	(2.37)
ΔInt_{t-1}	0.02	0.01	0.01	0.01	0.01
	(1.11)	(1.09)	(1.08)	(1.07)	(0.77)
$\Delta Beta_{t-1}$	−0.02***	−0.02***	−0.02***	−0.02***	−0.02***
	(−6.01)	(−5.98)	(−6.06)	(−6.05)	(−5.67)
ΔBTM_{t-1}	−0.02	−0.02	−0.02	−0.02	0.02
	(−1.37)	(−1.32)	(−1.31)	(−1.30)	(1.35)
$\Delta Log(mve)_{t-1}$	0.01**	0.01**	0.01**	0.01**	0.02***
	(2.04)	(1.97)	(2.03)	(2.01)	(5.23)
$\Delta Skew_{t-1}$	0.00	0.00	0.00	0.00	0.00
	(1.55)	(1.54)	(1.55)	(1.55)	(1.30)
$\Delta Turn_{t-1}$	−0.02	−0.02	−0.02	−0.02	−0.05
	(−0.27)	(−0.27)	(−0.26)	(−0.26)	(−0.84)
$ETFdummy_{t-1}$	−0.00	−0.00	−0.00	−0.00	−0.00
	(−0.05)	(−0.01)	(−0.17)	(−0.13)	(−0.42)
Constant	−0.01	−0.01	−0.01	−0.01	−0.02
	(−1.52)	(−1.48)	(−1.52)	(−1.48)	(−0.41)
Observations	28,390	28,390	28,390	28,390	28,390
R-squared	0.030	0.030	0.030	0.030	0.033
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel D: ΔVR(6)

	(1)	(2)	(3)	(4)	(5)
ΔETF_{SZCt-1}	15.91				10.64
	(0.89)				(0.63)
ΔETF_{SHCt-1}		−2.86**			−0.37**
		(−2.40)			(−2.32)
ΔETF_{SZSt-1}			−2.15*		−1.63***
			(−1.89)		(−3.16)
ΔETF_{SHSt-1}				−0.39**	−0.52**
				(−2.18)	(−2.20)
ΔInt_{t-1}	−0.02	−0.02	−0.02	−0.02	0.16***
	(−1.04)	(−1.09)	(−1.13)	(−1.15)	(8.93)
$\Delta Beta_{t-1}$	−0.01	−0.01	−0.01	−0.01	−0.01
	(−1.27)	(−1.24)	(−1.41)	(−1.42)	(−1.59)
ΔBTM_{t-1}	0.05***	0.05***	0.05***	0.05***	−0.00
	(3.04)	(3.21)	(3.22)	(3.27)	(−0.10)
$\Delta Log(mve)_{t-1}$	−0.08***	−0.08***	−0.08***	−0.08***	−0.14***
	(−18.47)	(−18.59)	(−18.48)	(−18.46)	(−22.70)
$\Delta Skew_{t-1}$	−0.00	−0.00	−0.00	−0.00	0.00
	(−0.76)	(−0.76)	(−0.75)	(−0.72)	(0.67)
$\Delta Turn_{t-1}$	−1.15***	−1.15***	−1.15***	−1.15***	−0.78***
	(−13.42)	(−13.42)	(−13.39)	(−13.40)	(−9.02)
$ETFdummy_{t-1}$	−0.01**	−0.01**	−0.01***	−0.01***	−0.00
	(−2.37)	(−2.29)	(−2.72)	(−2.64)	(−0.33)
Constant	0.03***	0.03***	0.03***	0.03***	0.03
	(2.59)	(2.72)	(2.59)	(2.72)	(0.42)
Observations	28,408	28,408	28,408	28,408	28,408
R-squared	0.070	0.069	0.069	0.069	0.076
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

Panel E: ΔVR(12)

(*continued on next page*)

Table 9 (*continued*)

Panel E: ΔVR(12)

	(1)	(2)	(3)	(4)	(5)
	(1)	(2)	(3)	(4)	(5)
ΔETF_{SZCt-1}	19.71				13.32
	(0.15)				(0.26)
ΔETF_{SHCt-1}		−3.72**			−1.83***
		(−2.14)			(−3.11)
ΔETF_{SZSt-1}			−4.16**		−3.95***
			(−2.50)		(−3.92)
ΔETF_{SHSt-1}				−1.12**	−1.80**
				(−2.35)	(−2.49)
ΔInt_{t-1}	−0.04	−0.04	−0.04	−0.04	0.22***
	(−1.50)	(−1.55)	(−1.56)	(−1.59)	(8.63)
$\Delta Beta_{t-1}$	0.01	0.01	0.01	0.01	0.01
	(1.50)	(1.53)	(1.40)	(1.38)	(1.09)
ΔBTM_{t-1}	0.08***	0.08***	0.08***	0.08***	−0.03
	(3.37)	(3.51)	(3.50)	(3.56)	(−1.08)
$\Delta Log(mve)_{t-1}$	−0.12***	−0.12***	−0.12***	−0.12***	−0.22***
	(−18.91)	(−19.02)	(−18.93)	(−18.92)	(−25.22)
$\Delta Skew_{t-1}$	−0.00	−0.00	−0.00	−0.00	0.00
	(−0.67)	(−0.67)	(−0.68)	(−0.63)	(0.88)
$\Delta Turn_{t-1}$	−1.54***	−1.54***	−1.53***	−1.53***	−0.97***
	(−12.24)	(−12.24)	(−12.20)	(−12.22)	(−7.73)
$ETFdummy_{t-1}$	−0.01**	−0.01**	−0.01**	−0.01**	0.00
	(−2.15)	(−2.06)	(−2.48)	(−2.37)	(0.11)
Constant	0.05**	0.05***	0.05**	0.05***	0.08
	(2.57)	(2.68)	(2.51)	(2.68)	(0.86)
Observations	28,408	28,408	28,408	28,408	28,408
R-squared	0.063	0.062	0.062	0.062	0.073
Year FE	YES	YES	YES	YES	YES
Industry FE	YES	YES	YES	YES	YES

This table presents the results of the regression analysis of changes in informational efficiency (i.e., *ΔNonsynch*, $\Delta LMSW_{dif}$, $\Delta LMSW_{res}$, *ΔVR(6)*, and *ΔVR(12)*) on changes in ETF ownership of different types (i.e., ΔETF_{SZC}, ΔETF_{SHC}, ΔETF_{SZS}, and ΔETF_{SHS}). Panels A–E are the corresponding results when the proxy for informational efficiency is changed to *ΔNonsynch*, $\Delta LMSW_{dif}$, $\Delta LMSW_{res}$, *ΔVR(6)*, and *ΔVR(12)*, respectively. t-statistics based on standard error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for variable definitions.

negative estimates of the coefficients of $\Delta Gamma_{i,\ t-1}$ all indicate that stock liquidity would attract more informed investors to the market, which may facilitate the informational efficiency of underlying stocks.

4.3.3. *Stock liquidity, informed trading, and informational efficiency*

As the last test of our channel analyses, we investigate whether the higher stock liquidity and greater participation by informed investors brought by ETFs contribute to the informational efficiency of its underlying stocks. We posit that there should be a positive relation between stock liquidity (informed investors) and price efficiency. To verify our conjecture, we run the following regression analysis:

$$\begin{aligned} \Delta Info_{it} &= \beta_0 + \beta_1 \Delta Liquidity_{it}(\Delta Informed_{it}) + \beta_2 \Delta Int_{it-1} + \beta_3 \Delta Beta_{it-1} \\ &+ \beta_4 \Delta BTM_{i,t-1} + \beta_5 \Delta Log\big(_{m} + \beta_6 \Delta Skew_{i,t-1} + \beta_7 \Delta Turn_{i,t-1} \\ &+ \beta_8 ETFdummy_{i,t-1} + \sum_j \beta_j Inddummy_j + \sum_l \beta_l Yeardummy + \varepsilon_{it} \end{aligned} \tag{24}$$

In Eq. (24), we regress changes in informational efficiency during year t on changes in stock liquidity or informed investors during the same period. Moreover, we control the same vector of control variables as in Eq. (14), and we add the year and industry-fixed effects.

Table 8 presents the regression analysis results. Panels A to E are the corresponding results when we measure informational efficiency using *Nonsynch*, $LMSW_{res}$, $LMSW_{dif}$, *VR(6)*, and *VR(12)*, respectively. For instance, Panel A shows that both changes in stock liquidity and changes in informed investors have a positive relation with changes in *Nonsynch*. The coefficients on changes in stock liquidity (Δ*Amivest*, Δ*Turn*, and Δ*Gamma*) and changes in informed investors (*ΔVPIN(8)* and *ΔVPIN(50)*) are all statistically significant and have the expected signs. The results in Panels B to E are quite similar. The findings are consistent with our conjecture that prices for component stocks become more efficient as ETFs improve stock liquidity and attract more informed investors.

These findings in channel analyses support the view that increased ETF ownership can lead to higher stock liquidity, which further attracts more informed traders. Consequently, both stock liquidity and participation by informed investors contribute to the pricing efficiency of underlying stocks. Our results provide empirical evidence for the price discovery theory in the market microstructure literature, which suggests that trading associated with the ETF-arbitrage mechanism can improve intraday price discovery for the underlying securities (Hasbrouck, 2003; Yu, 2005; Chen and Strother, 2008; Ivanov et al., 2013).

5. Further analyses

5.1. Implications of heterogeneity among the different types of ETF in conducting intraday trading

Different from Israeli et al. (2017), who find that ETF activity leads to the deterioration of pricing efficiency, we find that increases in ETF ownership are associated with increases in stock price informativeness. In nature, the contradicting results lie in the effect of ETFs on stock liquidity in different stock markets.

In the U.S. market, both the locking up of shares and the migration of uninformed investors away from underlying component securities would create a steady siphoning of firm-level liquidity, which generates a disincentive for informed traders to acquire information and thus decreases the informational efficiency of underlying securities. However, in the Chinese A-share market, which implements the "T + 1" trading restriction, the index arbitraging or intraday trading channel derives the positive relation between ETF ownership and stock liquidity. When investors conduct intraday trading using ETFs, the liquidity of underlying stocks increases, which in turn generates an incentive for informed traders to expand resources to obtain firm-specific information and thus increases the informational efficiency of underlying securities.

To corroborate this conjecture, we utilize the heterogeneity of four different types of ETFs in the Chinese market. As discussed in Section 2.1, ETFs in the Chinese A-share market can be grouped into four types: Shanghai single-market ETFs (ETF_{SHS}), Shenzhen single-market ETFs (ETF_{SZS}), Shanghai cross-market ETFs (ETF_{SHC}), and Shenzhen cross-market ETFs (ETF_{SZC}). Different from the first three types of ETFs, investors in ETF_{SZC} are unable to conduct intraday trading because of its low settlement efficiency. This heterogeneity of ETFs provides us with a unique setting to verify the underlying channels of our research and help us differentiate our work from piror studies, especially Israeli et al. (2017).

Table 10
ETF activity and contemporaneous return-earnings relation.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$
$Earn_{i,t}$	0.46*	0.50**	0.48**	0.48**	0.92***	0.90***	0.90***
	(2.00)	(2.99)	(2.76)	(2.85)	(3.22)	(3.05)	(3.41)
$\triangle ETF_{i,t}$	9.99**	10.81**		10.12**	12.13***	19.94***	18.13***
	(2.66)	(2.84)		(2.68)	(4.82)	(4.94)	(4.83)
$Earn_{i,t}*\triangle ETF_{i,t}$	35.53***	30.80**		9.10**	71.67**	74.98**	118.69*
	(3.03)	(2.93)		(2.24)	(2.76)	(2.95)	(1.99)
$Inst\_residual_{i,t}$		1.44***	1.46***	1.46***	1.28***	1.28***	1.26***
		(8.03)	(8.09)	(8.06)	(7.38)	(7.33)	(7.24)
$Earn_{i,t}*Inst\_residual_{i,t}$			−1.79	−1.04	−1.98	−1.62	−1.79
			(−1.23)	(−0.77)	(−1.36)	(−1.14)	(−1.17)
$MTB_{i,t-1}$					−0.02	−0.01	0.01
					(−0.34)	(−0.12)	(0.09)
$Size_{i,t-1}$					−0.07**	−0.09**	−0.10**
					(−2.21)	(−2.79)	(−2.87)
$STD_{i,t-1}$					−2.12**	−2.22**	−2.14**
					(−3.00)	(−3.00)	(−2.20)
$Ret_{i,(t-12,t-2)}$					−0.02	−0.01	−0.03
					(−0.63)	(−0.34)	(−0.69)
$Loss_{i,t}$					−0.00	−0.00	−0.01
					(−0.04)	(−0.20)	(−0.69)
$Earn_{i,t}* Loss_{i,t}$					0.04	0.10	0.12
					(0.28)	(0.70)	(1.02)
$Earn_{i,t-1}$						−0.08	−0.10
						(−0.84)	(−0.91)
$ETF_{i,t-1}$						9.97***	9.46***
						(4.85)	(5.07)
$Earn_{i,t-1}*\triangle ETF_{i,t}$							40.01
							(0.68)
$Beta_{i,t-1}$							−0.07
							(−1.51)
$Ret_{i,t+1}$							−0.18*
							(−2.07)
Constant	0.30	0.33	0.33	0.33	1.35**	1.69**	1.87**
	(1.54)	(1.72)	(1.72)	(1.72)	(2.40)	(2.82)	(3.00)
Adj.R^2	0.025	0.096	0.090	0.096	0.184	0.189	0.218
Observations	34,511	34,511	34,511	34,511	33,384	33,384	33,303

This table presents the replication results of Eq. (12) in Glosten et al. (2021), which examine the effect of ETF activity on contemporaneous return-earnings relation. t-statistics based on standard error clustered by the firm are shown in parentheses. ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively. See Appendix Table A.1 for main variable definitions.

According to the theories of Holden (1995) and Subrahmanyam and Titman (1999), if investors cannot use index-linked securities in conducting intraday arbitraging, then the ability to provide liquidity for underlying stocks is severely weakened. Our previous results also show that higher stock liquidity and more informed investor participation are driving forces in the relation between ETF ownership and stock price informativeness. As ETF_{SZC} cannot be used for conducting intraday trading, such as intraday arbitrage, its ability to provide liquidity and attract informed investors is greatly diminished. Therefore, if we obtain positive relations between all types of ETF ownership and pricing efficiency except for ETF_{SZC}, we could further verify the results of our channel tests.

We calculate the ownership of each type of ETF and examine the effects of their ownership on the informational efficiency of the underlying stocks. Table 9 presents the regression analysis results from the estimation of Eq. (14) by replacing our key independent variable with changes in ownership for the four different types of ETFs. Panels A to E are the corresponding results when we measure informational efficiency using *Nonsynch*, $LMSW_{res}$, $LMSW_{dif}$, *VR(6)*, and *VR(12)*, respectively. For instance, Panel A shows that the estimates for changes in Shenzhen cross-market ETF ownership (ΔETF_{SZC}) are negative and nonsignificant; however, estimates for changes in the other three types of ETF ownership—ΔETF_{SZS}, ΔETF_{SHC}, and ΔETF_{SHS}—are all positive and statistically significant. If we consider changes in ownership of all four different ETFs simultaneously in one regression, the significantly positive relation between ETF activities and pricing efficiency still only exists in the other three types of ETFs. The results in Panels B to D exhibit strong similarities to Panel A. These results not only help us rule out the potential reverse causality problem but also provide further evidence supporting our underlying mechanisms.

Table 11
ETF activity and contemporaneous return–earnings components relation.

	(1)	(2)	(3)	(4)	(5)
	Full	Small	Big	Low	High
	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$	$Ret_{i,t}$
$\triangle ETF_{i,t}$	19.40**	64.98**	9.30	39.79**	2.48
	(2.50)	(2.62)	(1.28)	(2.36)	(0.23)
$Earn\_Sys_{i,t}$	0.63	0.56	−2.95	0.18	−2.51
	(0.73)	(0.37)	(−1.29)	(0.13)	(−1.00)
$Earn\_Firm_{i,t}$	0.92***	1.16***	−0.23	1.13***	−1.15
	(3.94)	(4.42)	(−0.61)	(4.77)	(−1.39)
$Earn\_Sys_{i,t}*\triangle ETF_{i,t}$	−41.98	−411.05	67.46	−400.82*	47.74
	(−0.38)	(−1.57)	(0.48)	(−2.02)	(0.17)
$Earn\_Firm_{i,t}*\triangle ETF_{i,t}$	95.30**	123.13**	77.38**	76.96*	147.57**
	(2.73)	(2.66)	(2.42)	(2.14)	(2.16)
$Inst\_residual_{i,t}$	1.49***	1.55*	0.82**	1.93**	1.13***
	(3.75)	(2.14)	(2.16)	(2.61)	(4.80)
$Earn\_Firm_{i,t}*Inst\_residual_{i,t}$	−0.29	−0.49	−0.01	0.57	−3.48
	(−0.22)	(−0.25)	(−0.00)	(0.36)	(−1.41)
$Earn\_Sys_{i,t}*Inst\_residual_{i,t}$	−16.27**	−21.39*	−12.67**	−23.17*	−8.81**
	(−2.44)	(−1.83)	(−2.21)	(−2.06)	(−2.48)
$MTB_{i,t-1}$	0.01	0.05	0.01	0.07	−0.04
	(0.24)	(0.97)	(0.08)	(1.55)	(−0.37)
$Size_{i,t-1}$	−0.10**	−0.18***	−0.02	−0.14***	−0.05*
	(−2.85)	(−3.44)	(−1.16)	(−3.69)	(−2.01)
$STD_{i,t-1}$	−2.12**	−1.74*	−2.25	−1.76*	−1.06
	(−2.18)	(−1.78)	(−1.63)	(−1.79)	(−0.80)
$Ret_{i,(t-12,t-2)}$	−0.02	−0.03	0.02	−0.06	0.10
	(−0.57)	(−0.83)	(0.26)	(−1.74)	(1.54)
$Loss_{i,t}$	−0.01	−0.03***	0.11	−0.02**	0.12
	(−0.71)	(−3.27)	(1.42)	(−2.41)	(1.61)
$Earn_{i,t}* Loss_{i,t}$	0.05	−0.12***	2.94	−0.01	2.06**
	(0.75)	(−3.38)	(1.65)	(−0.10)	(2.28)
$Earn_{i,t-1}$	−0.09	−0.19	0.67**	−0.21	1.14*
	(−0.74)	(−1.37)	(2.72)	(−1.33)	(1.81)
$ETF_{i,t-1}$	9.57***	15.26***	3.57	13.90***	3.38**
	(5.07)	(3.79)	(1.58)	(4.08)	(2.38)
$Earn_{i,t-1}*\triangle ETF_{i,t}$	24.16	26.45	89.40	33.03	0.12
	(0.41)	(0.45)	(0.96)	(0.44)	(0.00)
$Beta_{i,t-1}$	−0.07	−0.07	−0.02	−0.08	−0.01
	(−1.53)	(−1.30)	(−0.38)	(−1.53)	(−0.26)
$Ret_{i,t+1}$	−0.18*	−0.20**	−0.12	−0.20**	−0.09
	(−2.04)	(−2.29)	(−1.11)	(−2.40)	(−1.16)
Constant	1.90**	3.21***	0.38	2.61***	0.88*
	(2.96)	(3.36)	(0.90)	(3.46)	(2.05)
Adj.R^2	0.223	0.220	0.285	0.217	0.316
Observations	33,303	24,898	8405	24,511	8792

5.2. ETF ownership and short-run information efficiency

In this section, we replicate the main empirical analyses in Glosten et al. (2021) and test whether and how ETF activity would improve short-run informational efficiency. Specifically, we estimate the following model as shown in Eq. (25) (i.e., the Eq. (12) in Glosten et al. (2021)):

$$\begin{aligned} Ret_{it} = \beta_0 + \beta_1 Earn_{i,t} + \beta_2 \Delta ETF_{i,t} + \beta_3 Earn_{i,t} \times \Delta ETF_{i,t} + \beta_4 Inst\_residual_{i,t} + \\ \beta_5 Earn_{i,t} * Inst\_residual_{i,t} + \beta_6 MTB_{i,t-1} + \beta_7 Size_{i,t-1} + \beta_8 STD_{i,t-1} + \\ \beta_9 Ret_{i,(t-12,t-2)} + \beta_{10} Loss_{i,t} + \beta_{11} Earn_{i,t} \times Loss_{i,t} + \beta_{12} Earn_{i,t-1} + \\ \beta_{13} ETF_{i,t-1} + \beta_{14} Earn_{i,t-1} \times \Delta ETF_{i,t-1} \beta_{15} \beta_{i,t-1} + \beta_{16} Ret_{i,t+1} + \varepsilon_{it} \end{aligned} \tag{25}$$

As shown in Table 10, we find that consistent with the findings of Glosten et al. (2021), the interaction between ETF activity and earnings information is statistically significant in explaining stock returns across different specifications, which suggests that ETF activity increases the return-earnings relation and thus improves the short-run informational efficiency of underlying stocks.

Second, we decompose $Earn_{i,t}$ into two components, systematic earnings news ($Earn\_Sys_{i,t}$) and firm-specific earnings news ($Earn\_Firm_{i,t}$), and regress $Ret_{i,t}$ on these two components and their interactions with $\Delta ETF_{i,t}$ by following Equaiton (26) (i.e., the Eq. (14) in Glosten et al. (2021)):

$$\begin{aligned} Ret_{it} = \beta_0 + \beta_1 \Delta ETF_{i,t} + \beta_2 Earn\_Sys_{i,t} + \beta_3 Earn\_Firm_{i,t} + \\ \beta_4 Earn\_Sys_{i,t} \times \Delta ETF_{i,t} + \beta_5 Earn\_Firm_{i,t} \times \Delta ETF_{i,t} + \\ \beta_6 Inst\_residual_{i,t} + \beta_7 Earn\_Sys_{i,t} * Inst\_residual_{i,t} + \\ \beta_8 Earn\_Firm_{i,t} * Inst\_residual_{i,t} + \beta_9 MTB_{i,t-1} + \beta_{10} Size_{i,t-1} + \\ \beta_{11} STD_{i,t-1} + \beta_{12} Ret_{i,(t-12,t-2)} + \beta_{13} Loss_{i,t} + \beta_{14} Earn_{i,t} \times Loss_{i,t} + \\ \beta_{15} Earn_{i,t-1} + \beta_{16} ETF_{i,t-1} + \beta_{17} Earn_{i,t-1} \times \Delta ETF_{i,t-1} + \beta_{18} \beta_{i,t-1} + \\ \beta_{19} Ret_{i,t+1} + \varepsilon_{it} \end{aligned} \tag{26}$$

Table 11 presents our results. The coefficient on the interaction of $Earn\_Firm_{i,t}$ and $\triangle ETF_{i,t}$ is positive and significant, implying that an increase in ETF activity pushes prices to reflect more firm-specific earnings news. The coefficient on the interaction of $Earn\_Sys_{i,t}$ and $\triangle ETF_{i,t}$ is insignificant, indicating that ETF activity does not increase short-run informational efficiency related to systematic fundamental information. The results are consistent for the full sample and different partitions. The evidence from Table 11 suggests that different from Glosten et al. (2021), the increase in short-run informational efficiency in the Chinese A-share market is attributable to contemporaneous firm-specific earning information rather than systematic accounting information.

5.3. Robustness tests

To verify the robustness of our results, we conduct the following tests. First, we change the proxies in our channel analyses. Following Goyenko et al. (2009) and Israeli et al. (2017), we use (1) the relative bid-ask spread ($HLSPREAD_{it}$) and (2) an adjusted measure of the price impact of trades ($ILLIQ\_N_{it}$) to capture stock liquidity. $HLSPREAD_{it}$ represents the annual high-low measure of the bid-ask spread for firm i over year t. A higher measure is associated with lower stock liquidity. $ILLIQ\_N_{it}$ denotes the numerator of the *Amivest* measure, which was constructed to mitigate the problem that ETF ownership can mechanically induce greater trading volume without conferring an overall improvement in the liquidity of the underlying stocks (Ben David et al., 2018). A higher value of $ILLIQ\_N_{it}$ is associated with a higher level of stock liquidity. In addition, for informed investors, we use institutional ownership as another proxy. After changing the above variables, we rerun our channel analysis tests and find results consistent with our earlier findings.

Second, our results are robust to alternative model specifications. For instance, we add a two-period lagged term of ETF ownership (ETF_{t-2}) as the right-hand-side variable, and the estimates of our key interest variable (ΔETF_{t-1}) are still statistically significant and with the expected signs. In addition, we use the heteroscedasticity-consistent standard errors clustered by firm and year to adjust our error terms, and the results stay quantitively similar.

Taken together, our results are robust to alternative proxies and model specifications, which verifies their robustness.

6. Conclusion

The Chinese A-share market has become the second-largest stock market in the world regarding market capitalization, and its unique features, such as the "T + 1" trading restriction, 10% daily price limits, dominance of retail investors, and short-sale constraints, distinguish it from developed markets. Given the importance and specificity of this market, how financial innovations (such as ETFs) affect the informational efficiency of its underlying stocks deserves further study.

Utilizing a comprehensive dataset of ETF ownership in the Chinese A-share market from 2006 to 2020, we examine the economic linkage between ETF ownership and the pricing efficiency of underlying stocks, which is of considerable interest to securities market regulators, financial practitioners, and academic researchers alike. Our main results show that increases in ETF ownership are associated with improvements in stock pricing efficiency, and our channel tests suggest that increased ETF ownership leads to higher stock liquidity and more participation by informed investors in stock trading, which results in higher informational efficiency of underlying stocks. To solve the endogeneity problem, we employ CSI 300 and CSI 500 index reconstitution as an exogenous shock to ETF ownership and exploit the instrumental variable (IV) method and corroborate our main findings. In addition, we exploit the unique features of ETFs in the Chinese market, utilize the heterogeneity of four types of ETFs and verify our underlying mechanism. Last but not least, we replicate the main empirical analyses in Glosten et al. (2021) and find that ETF activities contribute to the pricing efficiency of underlying securities also in the short run, and this can be attributed to its pricing discovery in firm-specific information. In sum, our results not only add to the debates on the consequences of ETFs but also shed light on the informational role of ETFs for emerging markets.

Appendix A. Appendix

Table A.1
Variable definitions.

Variable	Definition
Nonsynch	A logarithmic transformation of R^2 defined as log(1-R^2/R^2), where R^2 is estimated separately for each firm-year, as described in Section 3.1.1
$LMSW_{res}$	The dynamic volume–return relation obtained from the regression analysis of Eq. (4), as described in Section 3.1.1
$LMSW_{dif}$	The dynamic volume–return relation obtained from the regression analysis of Eq. (5), as described in Section 3.1.1
VR(12)	The variance ratio based on 12-period lags; the test-statistic is denoted in Eq. (7), as described in Section 3.1.1
VR(6)	The variance ratio based on 6-period lags; the test-statistic is denoted in Eq. (7), as described in Section 3.1.1
ETF	The proportion of shares owned by all candidate ETFs in a stock's total shares outstanding at the end of the year
ETF_{SZC}	The proportion of shares owned by Shenzhen cross-market ETFs in a stock's total shares outstanding at the end of the year
ETF_{SHC}	The proportion of shares owned by Shanghai cross-market ETFs in a stock's total shares outstanding at the end of the year
ETF_{SZS}	The proportion of shares owned by Shenzhen single-market ETFs in a stock's total shares outstanding at the end of the year
ETF_{SHS}	The proportion of shares owned by Shanghai single-market ETFs in a stock's total shares outstanding at the end of the year
Turn	The annual average of daily trading volume of a firm's share to its total market value
Amivest	The annual average of daily trading volume of a firm associated with a unit change in the stock price
Gamma	The coefficient obtained from the regression analysis of Eq. (11), as described in Section 3.1.3
VPIN(8)	The Volume-Synchronized Probability of Informed Trading (VPIN) metrix when n chooses 8 in Eq. (13), as described in Section 3.1.4.
VPIN (50)	The Volume-Synchronized Probability of Informed Trading (VPIN) metrix when n choose 50 in Eq. (13), as described in Section 3.1.4.
Ret	The annual return for a firm in a given year
Earn	Earnings before extraordinary items for a firm scaled by the market value of equity in a given year
EarnAgg	The fitted value from the firm-year estimation of the model in Eq. (18), as described in Section 4.1.2
EarnFirm	The residual value from the firm-year estimation of the model in Eq. (18), as described in Section 4.1.2
Int	The percentage of a firm's shares held by institutions at the end of the year
Beta	The coefficient obtained from the regression analysis of Eq. (15), as described in Section 4.1
BTM	The book-to-market ratio of a firm in a given year
Log(mve)	The natural logarithm of a firm's market value of equity at the end of the year
Skew	The skewness of a firm's daily return over a given year
AtGrowth	Total assets growth rate for a firm in two consecutive years
Stdret	The standard deviation of a firm's daily returns over year t
Intan	The ratio of intangible assets to total assets
Loss	An indicator variable that equals one if a firm experienced a loss (defined as Earn <0) in a given year, otherwise it equals zero
Mom	The cumulative stock return for months −12 to −1 relative to the year-end
Reconst	A dummy variable, which equals 1 if a stock moves from the CSI 300 to the CSI 500 in a given year and equals 0 if a stock moves from the CSI 500 to the CSI 300 in a given year.
Δ	The annual change of operator

This table displays definitions of variables used in our analyses.

References

Admati, A.R., Pfleiderer, P., 1988. A theory of intraday patterns: volume and price variability. Rev. Financ. Stud. 1 (1), 3–40.
Amihud, Y., Mendelson, H., Lauterbach, B., 1997. Market microstructure and securities values: evidence from the Tel Aviv stock exchange. J. Financ. Econ. 45 (3), 365–390.
Antoniou, C., Li, F.W., Liu, X., Subrahmanyam, A., Sun, C., 2020. The real effects of exchange-traded funds. Rev. Financ. Stud. forthcoming.
Ben David, I., Franzoni, F., Moussawi, R., 2018. Do ETFs increase volatility? J. Financ. 73 (6), 2471–2535.
Bond, P., Edmans, A., Goldstein, I., 2012. The real effects of financial markets. Ann. Rev. Finance Econom. 4 (1), 339–360.
Brennan, M.J., Schwartz, E.S., 1990. Arbitrage in stock index futures. J. Bus. 63 (1), S7–S31.
Carpenter, J.N., Lu, F., Whitelaw, R.F., 2021. The real value of China's stock market. J. Financ. Econ. 139 (3), 679–696.
Chen, G., Strother, T.S., 2008. On the contribution of index exchange traded funds to price discovery in the presence of price limits without short selling. Available at SSRN 1094485.
Chen, G.J., Zhang, R.Z., Xie, P.L., Zhao, X.Q., 2019. Informed trading, information uncertainty and stock return premium. Chin. J. Manag. Sci. 22 (4), 53–74 (in Chinese).
Collins, D.W., Kothari, S.P., Shanken, J., Sloan, R.G., 1994. Lack of timeliness and noise as explanations for the low contemporaneous return-earnings association. J. Account. Econ. 18 (3), 289–324.
Da, Z., Shive, S., 2018. Exchange traded funds and asset return correlations. Eur. Financ. Manag. 24 (1), 136–168.
Diamond, D.W., Verrecchia, R.E., 1981. Information aggregation in a noisy rational expectations economy. J. Financ. Econ. 9 (3), 221–235.
Durnev, A., Morck, R., Yeung, B., Zarowin, P., 2003. Does greater firm-specific return variation mean more or less informed stock pricing? J. Account. Res. 41 (5), 797–836.
Easley, D., Kiefer, N.M., O'Hara, M., Paperman, J.B., 1996. Liquidity, information, and infrequently traded stocks. J. Financ. 51 (4), 1405–1436.

Easley, D., López De Prado, M.M., O'Hara, M., 2012. Flow toxicity and liquidity in a high-frequency world. Rev. Financ. Stud. 25 (5), 1457–1493.
Fama, E.F., French, K.R., 1992. The cross-section of expected stock returns. J. Financ. 47 (2), 427–465.
Foster, F.D., Viswanathan, S., 1990. A theory of the interday variations in volume, variance, and trading costs in securities markets. Rev. Financ. Stud. 3 (4), 593–624.
Fremault, A., 1991. Stock index futures and index arbitrage in a rational expectations model. J. Bus. 523–547.
Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. J. Financ. Econ. 21 (1), 123–142.
Glosten, L., Nallareddy, S., Zou, Y., 2021. ETF activity and informational efficiency of underlying securities. Manag. Sci. 67 (1), 22–47.
Goyenko, R.Y., Holden, C.W., Trzcinka, C.A., 2009. Do liquidity measures measure liquidity? J. Financ. Econ. 92 (2), 153–181.
Grossman, S.J., Stiglitz, J.E., 1980. On the impossibility of informationally efficient markets. Am. Econ. Rev. 70 (3), 393–408.
Han, L.Y., Zheng, J.Y., Li, D.H., 2008. The feature of probability of informed trading and risk pricing in Shanghai stock market. Chin. J. Manag. Sci. 16 (1), 16–24 (in Chinese).
Hasbrouck, J., 2003. Intraday price formation in US equity index markets. J. Financ. 58 (6), 2375–2400.
Holden, C.W., 1995. Index arbitrage as cross-sectional market making. J. Futur. Mark. 15 (4), 423.
Huang, S., Hara, O., Zhong, Z.M., 2021. Innovation and informed trading: evidence from industry ETFs. Rev. Financ. Stud. 34 (3), 1280–1316.
Israeli, D., Lee, C., Sridharan, S.A., 2017. Is there a dark side to exchange traded funds? An information perspective. Rev. Acc. Stud. 22 (3), 1048–1083.
Ivanov, S.I., Jones, F.J., Zaima, J.K., 2013. Analysis of DJIA, S&P 500, S&P 400, NASDAQ 100 and Russell 2000 ETFs and their influence on price discovery. Glob. Financ. J. 24 (3), 171–187.
Jin, L., Myers, S.C., 2006. R2 around the world: new theory and new tests. J. Financ. Econ. 79 (2), 257–292.
Kumar, P., Seppi, D.J., 1994. Information and index arbitrage. J. Bus. 481–509.
Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica 1315–1335.
Lakonishok, J., Shleifer, A., Vishny, R.W., 1994. Contrarian investment, extrapolation, and risk. J. Financ. 49 (5), 1541–1578.
Lettau, M., Madhavan, A., 2018. Exchange-traded funds 101 for economists. J. Econ. Perspect. 32 (1), 135–154.
Li, B., Rajgopal, S., Venkatachalam, M., 2014. R2 and idiosyncratic risk are not interchangeable. Account. Rev. 89 (6), 2261–2295.
Llorente, G., Michaely, R., Saar, G., Wang, J., 2002. Dynamic volume-return relation of individual stocks. Rev. Financ. Stud. 15 (4), 1005–1047.
Lo, A.W., MacKinlay, A.C., 1988. Stock market prices do not follow random walks: evidence from a simple specification test. Rev. Financ. Stud. 1 (1), 41–66.
Pástor, L., Stambaugh, R.F., 2003. Liquidity risk and expected stock returns. J. Polit. Econ. 111 (3), 642–685.
Roll, R., 1984. A simple implicit measure of the effective bid-ask spread in an efficient market. J. Financ. 39 (4), 1127–1139.
Subrahmanyam, A., Titman, S., 1999. The going-public decision and the development of financial markets. J. Financ. 54 (3), 1045–1082.
Verrecchia, R.E., 1982. Information acquisition in a noisy rational expectations economy. Econometrica 1415–1430.
Wurgler, J., 2000. Financial markets and the allocation of capital. J. Financ. Econ. 58 (1–2), 187–214.
Yu, L., 2005. Basket securities, price formation, and informational efficiency. In: Price Formation, and Informational Efficiency (March 25, 2005).
Zhou, Q.L., Zhu, Y.J., Jia, L.X., 2015. Probability of informed trading, spuidity and volatility: evidence from China's stock index futures market. J. Financ. Res. 58 (5), 132–147 (in Chinese).