Finance Research Letters 62 (2024) 105108

Contents lists available at ScienceDirect

Finance Research Letters

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

ETF ownership and stock pricing efficiency: The role of ETF arbitrage

Guanhua Chen [a], Xiangli Liu [b], Xiao Liu [b], Zhihua Zhao [b,*]

[a] Law School, Beijing Technology And Business University, China
[b] School of Finance, Central University of Finance and Economics, China

ARTICLE INFO

JEL classification:
G12
G14
G181

Keywords:
ETF
Pricing efficiency
Arbitrage

ABSTRACT

This study investigates the effect of exchange-traded fund (ETF) ownership on the stock pricing efficiency in the Chinese market using a panel of stock-daily observations. Based on the ETF portfolio composition file (PCF) and the intraday trading data of stocks from 2012 to 2021, we estimate daily ETF holdings based on a new method and find that the increase in ETF ownership stimulates ETF arbitrage and intensifies the contagion of noise information from the ETF market to the stock market, reducing the pricing efficiency of the underlying stocks. We also employ the change of ETF trading policy from the Shenzhen Stock Exchange in 2019 as a quasi-natural experiment to alleviate the endogeneity issue. Our study proposes a new method to estimate ETF holdings and provides new evidence on the financial consequences of ETFs.

1. Introduction

As an important financial innovation, exchange-traded funds (ETFs) facilitate passive investment in the underlying indexes at low cost and high liquidity, which have grown rapidly over the past two decades, especially in China. Fig. 1 illustrates the changes in the number of ETFs and AUM (assets under management) in China. By the end of September 2023, there were 853 equity-typed ETFs in China, managing approximately 2 trillion yuan in assets, but still far below the $8 trillion in the US ETF market.

The uniqueness of ETFs lies in their dual-market structure: they can be traded in the ETF secondary market or subscribed and redeemed in the ETF primary market, which provides new arbitrage opportunities. Theoretically, the ETF arbitrage mechanism can promote price discovery (Box et al., 2021). Some studies support the positive role of ETFs in pricing efficiency and stock liquidity (Huang et al., 2021; Glosten et al., 2021; Wu and Zhu, 2023). However, with the rise of ETFs, their negative impacts have raised significant concerns. The SEC's analysis of the May 6, 2010, Flash Crash in the United States attributes the accelerated market downturn to the arbitrage activities between ETFs and the underlying stocks (Aldridge, 2016). A similar observation was made during the 2015 Chinese stock market crash when ETF trading volume notably surged. Recent literature also finds that ETFs increase stock volatility and return comovement and decrease stock pricing efficiency and liquidity (Israeli et al., 2017; Da and Shive, 2018; Ben-David et al., 2018).

These seemingly contradictory findings drive an interesting question: how does ETF ownership affect the pricing efficiency of the underlying stocks in the Chinese market? Given the immaturity of the Chinese ETF market, its impact on the stock market remains unclear. The stock price is a critical information aggregator and indicator, and its pricing efficiency is crucial for the capital market and

* Corresponding author.
E-mail address: zzh@email.cufe.edu.cn (Z. Zhao).

https://doi.org/10.1016/j.frl.2024.105108
Received 27 November 2023; Received in revised form 28 January 2024; Accepted 5 February 2024
Available online 6 February 2024
1544-6123/

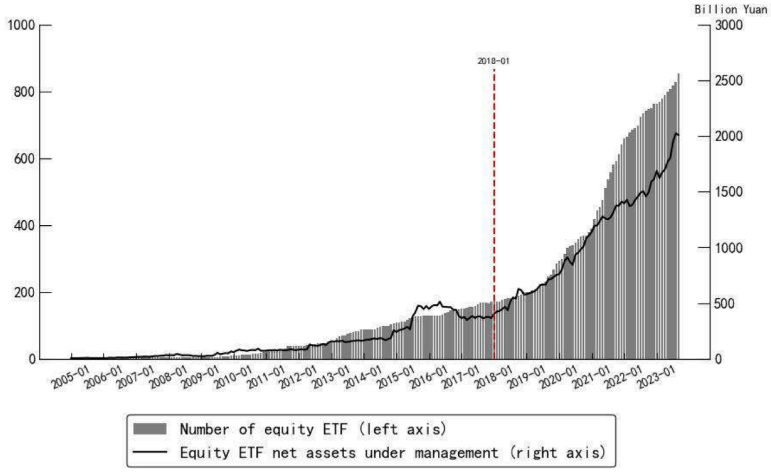

Fig. 1. The number and AUM of ETFs in China (2005–2023). Source: Wind database.

real economy (Goldstein, 2023). Therefore, exploring this issue can help us understand the financial consequences of the Chinese ETF market.

In this paper, we employ a new method and detailed data to investigate the effect of ETF ownership on the pricing efficiency of the underlying stocks in the Chinese market. Specifically, to address the limitations of prior studies that estimate daily ETF holdings using ETF annual reports, we calculate the daily change in ETF ownership based on the daily ETF portfolio composition file (PCF). In addition, we use 1 min intraday high-frequency trading data of stocks to compute pricing efficiency measures, which can capture more details of trading information. Using the data from 2012 to 2021 in the Chinese market, we find that the increase in ETF ownership will reduce the pricing efficiency of the underlying stocks. We further demonstrate that ETF arbitrage is the key channel of this effect, which causes the contagion of noisy information from the ETF market to the stock market. To alleviate the endogeneity problem, we also employ the ETF subscription and redemption policy change from the Shenzhen Stock Exchange in 2019 as a quasi-natural experiment and construct an instrumental variable for testing, which still supports our hypotheses. Last but not least, we replicate prior empirical work, compare our results with theirs, and attribute the differences to the more granular data and method we use.

We contribute in three ways: First, we propose a new method with higher frequency to estimate daily ETF ownership, which can capture the daily information of ETF activities. Second, we provide new empirical evidence that ETF activities decrease the pricing efficiency of the underlying stocks, which is different from previous literature that highlights the positive effect of ETF ownership on pricing efficiency (Glosten et al., 2021; Wu and Zhu, 2023). Thirdly, we contribute to the studies on arbitrage activities in the Chinese market. We use the 2019 ETF policy change as an exogenous shock to overcome the potential endogeneity problem, which provides a new test scenario to examine the impact of arbitrage activities.

The remainder of this paper is organized as follows: Section 2 introduces the literature review and hypothesis development. Section 3 describes our data and variable construction method. Section 4 presents our empirical results. Section 5 concludes.

2. Literature review and hypothesis development

Extant studies hold different opinions on the effects of ETFs on the underlying securities markets (Ben-David et al., 2017). On the one hand, some researchers argue that ETFs have positive impacts on the price efficiency and liquidity of the underlying stocks due to ETFs' low cost, high liquidity, unique arbitrage mechanism, and greater public scrutiny (Marshall et al., 2013; Boone and White, 2015; Apple et al., 2016; Glosten et al., 2021).

On the other hand, recent evidence suggests that ETFs lead to negative financial consequences for the underlying asset. Pan and Zeng (2017) and Dannhauser (2017) find that ETFs decrease the liquidity of the underlying stocks due to the crowding-out effect or the complementary effect. In addition, Israeli et al. (2017) reveal the dark side of ETFs from an information perspective. They believe that ETF ownership increases trading costs and inhibits the incentive of informed investors to acquire stock-specific information, ultimately leading to a decrease in the price informativeness of the underlying assets. Ben-David et al. (2018) demonstrate that ETFs increase the volatility of the underlying assets because they attract more short-term noise traders to inject noise information into prices. Da and Shive (2018) show that ETFs contribute to excess return comovement due to faster incorporation of common information rather than idiosyncratic information. All these studies support the argument that ETFs harm the pricing efficiency of the underlying stocks.

ETF arbitrage serves as a key link between the ETF market and the underlying securities market. Authorized participants will narrow the spread between the two assets by creating or redeeming ETF shares (Box et al., 2021). However, ETF arbitrage activities also propagate non-fundamental demand shocks from ETFs to the underlying securities (Brown et al., 2021).

Prior literature focuses on the US ETF market, and there are few studies on the Chinese ETF market. However, there are some differences between the U.S. and Chinese ETF markets. Chinese ETF holders are predominantly retail investors, which differs significantly from the predominance of institutional investors in the United States. Meanwhile, redemption transactions in the primary market of Chinese ETFs are not restricted to authorized participants (APs). Moreover, investors are subject to "$T+1$", limit system and other trading restrictions on stock trading in China. Thus, compared with the mature US ETF market, we argue that the Chinese ETF market is relatively imperfect, where the information in ETF price is mainly market information and noise information, with less stock idiosyncratic information. We expect that in China, an increase in ETF ownership will transmit the non-firm-specific information from the ETF market to the stock market through ETF arbitrage activities, leading to a decrease in the pricing efficiency of the underlying stocks. Therefore, we propose two hypotheses:

Hypothesis 1. *The increase in ETF ownership will reduce the pricing efficiency of the underlying stocks.*

Hypothesis 2. *ETF arbitrage is the key channel through which ETF ownership decreases the pricing efficiency of the underlying stocks.*

3. Data and empirical model

3.1. Data source and sample selection

The sample period for this study spans from 2012 to 2021. We first collect data on all equity-typed ETFs traded in the Chinese A-share market from the China Securities Market and Accounting Research (CSMAR) database and the Wind database. Specifically, our dataset includes quarterly ETF holdings reports, daily ETF portfolio composition files (PCF), daily ETF shares, and announcements of splits and mergers. We exclude the samples of ETFs that: 1) have been established for less than six months; 2) are delisted during the sample period; 3) contain stocks that are not listed in the A-share market. Next, we also obtain the daily and 1 min intraday high-frequency trading data of all stocks in the Chinese stock market during the same period from the CSMAR database and the RESSET high-frequency database. We exclude the stock samples that: 1) receive special treatment (including ST, *ST, PT); 2) belong to the financial industry; 3) had been listed for less than six months.

3.2. Pricing efficiency

3.2.1. Pricing error

Following prior literature (Hasbrouck, 1993; Xiong et al., 2017), we exploit *Pricing Error (PE)* as an inverse proxy for stock pricing efficiency. Hasbrouck (1993) decomposed the intraday stock price into two parts: the random walk process based on the stock's fundamental value and the stationary residual caused by noise trading. A lower proportion of the residuals indicates higher stock pricing efficiency. By setting the stock price return and trade sign indicator as the endogenous variables, we can estimate a VAR model and obtain the variance of the residual series. Finally, we compute the ratio of the variance of intraday residual σ_s^2 to the variance of intraday return σ_r^2, which is defined as:

$$PE2_{i,d} = \frac{\sigma_s^2}{\sigma_r^2} \tag{1}$$

In addition, we also follow the method proposed by Rosch et al. (2017) that set four variables including price return, trade sign indicator, the signed volume and the sign of the trade times the square root of the number of shares traded as the endogenous variables to obtain a proxy variable *PE4* for the robustness test.

3.2.2. Variance ratio

Following Lo & MacKinlay (1988), we compute the *Variance Ratio (VR)* as an alternative proxy for pricing efficiency, which is defined as:

$$\mathrm{VR}\ (\mathrm{M,N}) = \left| \frac{\frac{\mathrm{Var}(p_t - p_{t-M})}{M}}{\frac{\mathrm{Var}(p_t - p_{t-N})}{N}} - 1 \right|, \tag{2}$$

where we set M equal to 5 or 30 and set N equal to 1, and denote them as *VR (5) and VR (30),* respectively.

3.2.3. Information share

Referring to Brogaard et al. (2022), we use the intraday trading data to calculate the proportion of four types of information in stock price: noise information, market information, private and public information, which are defined as:

$$\begin{aligned} Market_{i,t} &= \frac{\theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2}{(\sigma_w^2 + \sigma_s^2)} \\ Private_{i,t} &= \frac{\theta_x^2 \sigma_{\varepsilon_x}^2}{(\sigma_w^2 + \sigma_s^2)} \\ Public_{i,t} &= \frac{\theta_r^2 \sigma_{\varepsilon_r}^2}{(\sigma_w^2 + \sigma_s^2)} \\ Noise_{i,t} &= \frac{\sigma_s^2}{(\sigma_w^2 + \sigma_s^2)}. \end{aligned}, \quad (3)$$

where σ_w^2 denotes the total effective variance of stock price and is equivalent to $\sigma_w^2 = \theta_{r_m}^2 \sigma_{\varepsilon_{r_m}}^2 + \theta_x^2 \sigma_{\varepsilon_x}^2 + \theta_r^2 \sigma_{\varepsilon_r}^2$, σ_s^2 denotes the variance of noise, $\sigma_{\varepsilon_{rm}}^2$, $\sigma_{\varepsilon_x}^2$, $\sigma_{\varepsilon_r}^2$ denotes the variance of market information, private information, and public information, and θ_{r_m}, θ_x, θ_r denotes the coefficients of these three kinds of information in the stock price regression equation, respectively.

3.3. ETF ownership

ETF ownership is the proportion of shares owned by ETFs in a stock's total outstanding shares, which can be viewed as a proxy for the ETF activity corresponding to each stock. Prior literature usually computes ETF ownership for a stock based on the latest ETF holdings reports and assumes that daily ETF ownership remains unchanged during the non-reporting period, which is somewhat not rigorous and crude due to daily stock price fluctuations and portfolio adjustment. In this paper, we employ a more granular approach to estimating daily ETF ownership for each stock. We exploit the daily updated ETF portfolio composition file (PCF), also known as the subscription and redemption list, which lists a basket of the underlying stocks required to subscribe new ETF shares on that day, reflecting the daily adjustment of ETF holdings by an ETF manager who tracks the benchmark index. For example, on the first day, PCF requires 100 shares of stock A and 200 shares of stock B to subscribe for 10,000 shares of ETF J. Then, on the second day, we observe the net shares of ETF J increase by 10,000 shares, and we can expect that the shares of stock A and stock B in the portfolio of ETF J increase by 100 and 200 shares, respectively. The underlying index of the ETF will be adjusted every six months. In order to solve the problem of calculation error caused by the index adjustment, we set 14 June and 15 December as the index adjustment date of each year (we can observe significant changes in the PCF for those dates), and assume that the difference of the data of the methodology at the reporting time point is caused by the change of the underlying index. Then we spread the difference evenly during this period. Thus, we iterate this method at the ETF level from the ETF listing day to the next semi-annual report date or annual report date, and then take the new ETF holdings report as the new starting point for the next round of iterative calculation. Finally, we aggregate daily holdings of all ETFs and compute the daily ETF ownership at the stock level, which is defined as:

$$\text{ETF\_ratio}_{i,t} = \frac{\text{ETF share}_{i,t}}{Mkt\ share_{i,t}} = \frac{\sum_{j,t}^{J} ETFholding_{j,i,t}}{Mkt\ share_{i,t}}, \quad (4)$$

where $Mktshare_{i,t}$ represents the outstanding share of stock i on the day t, $ETFholdingshare_{i,t}$ represents the share of stock i held by all ETFs, and J represents the number of ETFs holding stock i.

3.4. ETF arbitrage

ETFs' unique subscription and redemption mechanism provides space for arbitrage activities, allowing the shock transmission from the ETF market to the stock market. According to prior studies (Ben-David et al., 2018), daily ETF arbitrage activities for a stock can be proxied by the weighted average value of net share changes in the ETF primary market, which is defined as:

$$\text{Flow}_{j,t} = abs\left(\frac{\text{Share}_{j,t} - \text{Share}_{j,t-1} \times Change_{j,t}}{\text{Share}_{j,t-1}}\right), \quad (5)$$

$$\text{Flow}_{i,t} = \sum_{j}^{J} Flow_{j,t} * \text{w}_{j,i,t-1}, \quad (6)$$

where $Share_{j,t}$ is the daily share of ETF j on the day t, $Change_{j,t}$ reflects the share change of ETF j caused by dividend, splitting, and other activities on the day t, $Flow_{j,t}$ is the ETF-level proxy, reflecting the percentage of share change of ETF j in the ETF primary market, $Flow_{i,t}$ is the stock-level proxy equal to the weighted sum of the ETF-level values, with the weights $\text{w}_{j,i,\ t-1}$ based on the proportion of stock i held in the ETF j.

3.5. Empirical model

To test the impact of ETF ownership on the pricing efficiency of the underlying stocks, we construct the following regression model:

Table 1
Definition and descriptive statistics.

Variable categories	Variable symbol	Definition	Obs.	Mean	Std	Min	Max
Independent Variable	ETF_ratio	ETF ownership see 3.3	5,945,306	0.237	0.487	0	2.689
Dependent Variable:	PE2	Pricing error see3.2	5,631,433	0.024	0.0439	0.000	0.273
Pricing Efficiency	PE4	Pricing error see3.2	5,631,433	0.028	0.0513	0.000	0.321
	VR(5)	Variance ratio see3.2	5,919,801	0.626	0.643	0.005	5.877
	VR(30)	Variance ratio see3.2	5,918,933	0.330	0.226	0.000	1.143
Dependent Variable:	Market	Market information share see 3.2	5,909,457	0.113	0.109	0.000	0.438
Price Information	Private	Private information share see 3.2	5,909,457	0.035	0.052	0	0.282
Share	Public	Public information share see 3.2	5,909,457	0.527	0.231	0.043	0.928
	Noise	Noise information share see 3.2	5,909,457	0.324	0.194	0.022	0.842
Arbitrage Activity	Flow	ETF arbitrage activity see 3.4	5,945,306	0.4060	1.0342	0	5.879
Control variable	Size	Logarithm of market value of stock shares outstanding	5,945,306	15.290	1.108	13.092	18.432
	Beta	Systemic risk of stocks	5,397,226	1.062	0.283	0.342	1.763
	Price	The inverse of the stock closing price	5,945,306	0.103	0.077	0.009	0.382
	BM	Book-to-market ratio of the stock	5,945,306	0.429	0.310	0.045	1.632
	Trade	Logarithm of stock trading amount	5,945,306	18.061	1.288	15.245	21.266
	Rv	Stock realized volatility	5,945,306	8.700	10.480	0.707	61.42

This table presents the summary statistics of variables used in our empirical study. The column labelled Obs, Max, Min, Mean and Std denote the observation, maximum, minimum, mean and Standard deviation values of those variables respectively.

$$PriceEff_{i,t} = \beta_0 + \beta_1 * ETF\_ratio_{\ i,t-1} + \sum_{n\geq 2} \beta_n * X_{n,i,t} + \theta_i + \gamma_t + \varepsilon_{i,t} \tag{7}$$

where the stock pricing efficiency $PriceEff_{i,t}$ is the dependent variable denoted by *PE*2, *VR(5),* and four types of information shares; the independent variable $ETF\_ratio_{\ i,t-1}$ is the one-period lagged proportion of stock*i* held in all ETFs; $X_{n,i,\ t}$ are the control variables; θ_i and θ_i denote the firm fixed effect and time fixed effect, respectively. Following Xiong et al. (2017), we control the stock trading characteristics, including market value $Size_{i,t}$, book-to-market ratio $BM_{i,t}$, the inverse of the closing price $Price_{i,t}$, the systemic risk indicator $Beta_{i,t}$, stock trading amount $Trade_{i,t}$ and stock realized volatility $Rv_{i,t}$.

3.6. Descriptive statistics

Table 1 presents the descriptive statistics for our main variables. All continuous variables are winsorized at the 1 % and 99 % quantiles to reduce the influence of outliers. The mean value of the independent variable *ETF_ratio* is 0.237 %, which is much smaller than 4.4 % for the United States market (Antoniou et al., 2023). Among the four kinds of information shares, public information predominates in stock prices with an average of 52.7 %, followed by noise information at 32.4 %, surpassing market information at 11.3 %. Private information contributes the least, averaging just 3.5 %.

4. Empirical results

4.1. Baseline regression

We first examine the effect of ETF ownership on the pricing efficiency of the underlying stocks. Table 2 shows the regression results of Eq. (7). Columns (1)–(3) reveal that pricing error exhibit a significant positive relation with *ETF_ratio*. Similar results are presented in columns (4)–(6), with a significantly positive effect of ETF ownership on the variance ratio. For four types of information shares, the results in columns (7)–(10) suggest that ETF ownership will increase the proportion of noise information and market information in the underlying stock prices, and reduce the share of firm's public information, but have no significant impact on the share of firm's private information. We also observe that noise information is most sensitive to changes in ETF ownership. Our results remain robust after adding different fixed effects or lagged terms of the dependent variable. Overall, the baseline results show that the increase in ETF ownership will reduce the pricing efficiency of the underlying stocks, supporting our *Hypothesis 1*.

4.2. Channel test

In this section, we further explore the channel through which ETF activities decrease the pricing efficiency of the underlying stocks. To investigate the role of ETF arbitrage, we estimate the following regression analysis:

$$Flow_{i,t} = \alpha_0 + \alpha_1 * ETF\_ratio_{\ i,t-1} + \sum_{n} \delta_n * X_{n,i,t} + \theta_i + \gamma_t + \varepsilon_{i,t}, \tag{8}$$

$$PriceEff_{i,t} = \beta_0 + \beta_1 * ETF\_ratio_{\ i,t-1} + \beta_2 * Flow_{i,t} + \sum_{n} \delta_n * X_{n,i,t} + \theta_i + \gamma_t + \varepsilon_{i,t}, \tag{9}$$

Table 2
ETF ownership and pricing efficiency.

	PE2			VR(5)			Noise	Market	Private	Public
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
ETF_ratio (t-1)	0.0053***	0.0030***	0.0024***	0.0625***	0.0418***	0.0344***	0.0101***	0.0070***	-0.0002	-0.0186***
	(13.49)	(8.73)	(9.29)	(13.88)	(10.08)	(9.94)	(11.20)	(12.10)	(-0.99)	(-17.84)
Beta	-0.0075***	-0.0061***	-0.0036***	-0.1146***	-0.0891***	-0.0719***	0.0061***	0.0383***	-0.0063***	-0.0437***
	(-14.20)	(-13.69)	(-11.14)	(-17.77)	(-16.06)	(-15.75)	(4.52)	(53.96)	(-25.46)	(-27.36)
Price	0.2535***	0.2256***	0.1618***	3.1549***	2.9430***	2.4137***	0.7852***	-0.0822***	-0.0228***	-0.6745***
	(42.02)	(33.83)	(35.97)	(50.90)	(42.29)	(44.08)	(59.00)	(-14.21)	(-14.29)	(-52.22)
BM	0.0030***	0.0036***	0.0011	0.0683***	0.0633***	0.0472***	0.0197***	-0.0011	0.0029***	-0.0218***
	(2.79)	(2.70)	(1.13)	(5.90)	(4.71)	(4.21)	(6.73)	(-0.75)	(6.85)	(-6.97)
Size	0.0098***	0.0088***	0.0073***	0.0858***	0.0782***	0.0653***	0.0194***	0.0014***	-0.0027***	-0.0186***
	(33.93)	(24.44)	(27.49)	(26.42)	(19.82)	(20.21)	(22.15)	(2.97)	(-18.33)	(-19.39)
Trade	-0.0148***	-0.0168***	-0.0139***	-0.1304***	-0.1448***	-0.1168***	-0.0382***	0.0126***	0.0037***	0.0211***
	(-59.55)	(-63.43)	(-69.74)	(-55.00)	(-64.19)	(-64.25)	(-73.29)	(64.03)	(42.84)	(39.31)
Rv	0.0004***	0.0005***	0.0004***	-0.0080***	-0.0078***	-0.0082***	-0.0005***	-0.0019***	0.0010***	0.0015***
	(33.00)	(43.44)	(45.00)	(-48.84)	(-56.18)	(-63.18)	(-13.93)	(-116.46)	(112.05)	(37.18)
lag(y)			0.2671***			0.1790***	0.1384***	0.2491***	0.0149***	0.1418***
			(66.07)			(75.70)	(84.23)	(136.03)	(23.28)	(111.45)
Constant	0.1158***	0.1682***	0.1376***	1.4714***	1.8515***	1.4811***	0.5750***	-0.1789***	0.0084***	0.4731***
	(30.79)	(29.07)	(30.44)	(32.61)	(30.64)	(28.67)	(40.71)	(-24.00)	(3.48)	(30.57)
Firm fixed effect	NO	YES	YES	NO	YES	YES	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,136,608	5,136,605	3,890,334	5,385,496	5,385,493	4,233,752	3,994,876	3,994,876	3,994,876	3,994,876
Adjusted R2	0.399	0.444	0.491	0.342	0.369	0.396	0.337	0.555	0.062	0.338

This table reports the baseline regression results of Eq. (7). PE2 and VR(5) are price efficiency variable, where Noise, Market, Private and Public represents noise information share, market information share, private information share and public information share in stock price respectively. The definitions of other control variables are provided in Table 1. Values in parentheses are the *t*-values,***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 3
Channel test.

	Flow	PE2		VR(5)		Noise	Market	Private	Public
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
ETF_ratio (t-1)	1.6609***		0.0017***		0.0227***				
	(66.75)		(4.50)		(5.01)				
Flow		0.0011***	0.0006***	0.0142***	0.0073***	0.0040***	0.0021***	-0.0000	-0.0066***
		(8.87)	(3.95)	(11.20)	(4.76)	(12.30)	(9.90)	(-0.11)	(-17.00)
Control	YES	YES	YES	YES	YES	YES	YES	YES	YES
Firm fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	5,392,021	3,890,334	3,890,334	4,233,752	4,233,752	3,994,876	3,994,876	3,994,876	3,994,876
Adjusted R2	0.744	0.482	0.482	0.405	0.405	0.337	0.554	0.062	0.338

This table mainly reports the channel test results of Eqs. (8) and (9). Flow represents the arbitrage activities, and the definitions of other variables are provided in Table 1. Columns (1) shows the relationships between ETF ownership and arbitrage activities. Columns (2), (4) and (6)–(9) show the results of Eq. (8) where columns (3) and (5) show the results of Eq. (9). Values in parentheses are the *t*-values,***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 4
The IV-2SLS model of ETF arbitrage.

	Flow (1)	PE2 (2)	Flow (3)	VR(5) (3)
Treat * Post	0.0565***		0.0551***	
	(17.86)			
$\widehat{Flow_{i,t}}$		0.0680***		0.0573*
		(14.82)		(1.81)
Control	YES	YES	YES	YES
Firm fixed effect	YES	YES	YES	YES
Time fixed effect	YES	YES	YES	YES
Observations	1,339,731	1,339,731	1,399,626	1,399,626

This table reports the results of IV-2SLS model during the period of October 21, 2018 to October 21, 2020. Treat represents whether a stock is held in a cross-market ETF in the Shenzhen Stock Exchange and Post represents whether the trade day is after October 21, 2019. The definitions of other variables are provided in Table 1. Columns (1) and (3) show the first stage result while columns (2) and (4) show the second stage results. Values in parentheses are the t-values, ***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

Table 3 presents the results of the channel test. Column (1) shows that the increase in ETF ownership will lead to stronger arbitrage activities. Columns (2)–(5) further reveal the negative impact of arbitrage activities on the pricing efficiency of the underlying stocks. In addition, the results of the four types of information shares in columns (6)–(9) show that ETF arbitrage increases the proportion of market information and noise information, and reduces the ratio of stock-specific information in the stock price. These results show that ETF arbitrage is the channel through which ETF ownership affects the pricing efficiency of underlying stocks, which supports our *Hypothesis 2*.

4.3. Endogeneity treatment

A potential reverse causality relationship exists between ETF arbitrage activity and the underlying stocks' pricing efficiency. In theory, when a stock is priced inefficiently, arbitrageurs are more likely to use the price difference between the two markets to realize arbitrage gains. To address this endogeneity problem, we use the ETF subscription and redemption policy change from the Shenzhen Stock Exchange in 2019 as a quasi-natural experiment. Before 2019, the Shenzhen Stock Exchange prohibited cross-market ETF shares subscribed in the ETF primary market from being sold in the ETF secondary market on the same day, which limited premium arbitrage activities. However, on September 27, 2019, this trading restriction was canceled,[1] which provided us with a good exogenous shock to arbitrage activities, because this new rule can stimulate arbitrage activities and is unlikely to be caused by factors at the individual stock level.

Then, we construct an instrumental variable of ETF arbitrage intensity and conduct the IV-2SLS regression. Specifically, we focus the sample in the year before and after the policy change and set dummy variables $Post_{i,t}$ and $Treat_{i,t}$. When the trading day is after October 21, 2019, $Post_{i,t}$ is equal to 1, and 0 otherwise. If a stock is held in a cross-market ETF in the Shenzhen Stock Exchange, $Treat_{i,t}$ is equal to 1, and 0 otherwise. First, we conduct a first-stage regression on the intensity of ETF arbitrage activity using the interaction term of $Post_{i,t}$ and $Treat_{i,t}$. Further, we use the fitted value of arbitrage intensity obtained from the first-stage regression to conduct the

[1] The Shenzhen Stock Exchange announced the policy change on September 27, 2019 and officially implemented the new policy on October 21, 2019.

Table 5
Comparison with prior literature.

	$\Delta PE\_2_{i,Y}$		$\Delta VR(5)_{i,Y}$	
	(1)	(2)	(3)	(4)
ΔETF_ratio $_{i,Y-1}$	0.0017***	0.0019***	0.0634***	0.0693***
	(5.14)	(5.23)	(11.66)	(11.37)
$\Delta Beta_{i,Y}$	-0.0085***	-0.0092***	-0.1057***	-0.1108***
	(-21.01)	(-21.84)	(-11.46)	(-11.43)
$\Delta Price_{i,Y}$	0.2407***	0.2551***	3.0168***	3.0973***
	(41.82)	(40.86)	(35.13)	(30.91)
$\Delta BM_{i,Y}$	0.0081***	0.0066***	0.1622***	0.1694***
	(7.07)	(5.61)	(9.06)	(8.18)
$\Delta Size_{i,Y}$	0.0067***	0.0106***	0.0821***	0.1109***
	(19.13)	(23.47)	(12.50)	(13.03)
$\Delta Trade_{i,Y}$	-0.0094***	-0.0101***	-0.1408***	-0.1462***
	(-39.17)	(-39.90)	(-24.97)	(-24.88)
$\Delta Rv_{i,Y}$	0.0004***	0.0004***	0.0091***	0.0090***
	(15.25)	(13.90)	(8.12)	(7.84)
$\Delta Ins\_ratio_{i,Y}$	-0.0001***	-0.0001***	-0.0007***	-0.0008***
	(-7.93)	(-9.29)	(-3.62)	(-4.00)
Constant	0.0003***	-0.0003***	0.0039**	-0.0005
	(3.02)	(-3.56)	(2.23)	(-0.25)
Firm fixed effect	NO	YES	NO	YES
Time fixed effect	YES	YES	YES	YES
Observations	20,411	20,293	20,649	20,531
Adjusted R2	0.539	0.591	0.354	0.377

This table reports the results of replicating Wu and Zhu (2023) methodology but using the data from our paper averaged over years. We aggregate our daily data at the annual level and calculate the annual change values and the raw definitions of variables are similar to baseline regression which are shown in Table 1. Values in parentheses are the *t*-values,***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

second-stage regression on the pricing efficiency of the underlying stocks.

The two-stage results are presented in Table 4. The results of the first-stage regression in columns (1) and (3) show that the interaction term coefficient is positive and significant at the 1 % level, implying that the policy shock in 2019 significantly increases the intensity of ETF arbitrage activities. The coefficients of the second-stage regression in columns (2) and (4) are still significantly positive, indicating that the increase in ETF arbitrage activities reduces the pricing efficiency of the underlying stocks. Overall, the results of our IV-2SLS model further confirm the channel role of ETF arbitrage and support our *Hypothesis 2*.

4.4. Robustness test

To verify the robustness of our conclusions, we first contrast our results with prior literature, and then we conduct other robustness checks.

First, we replicate the empirical work in Wu and Zhu (2023), where they employ the change of year-end ETF ownership as the proxy for ETF activity and use the change of annual aggregated values of daily trading data to compute pricing efficiency. Similarly, we aggregate our daily data at the annual level and calculate the annual change values. The new results in Table 5 still support our conclusion, which is different from Wu and Zhu (2023). We attribute the difference to our more granular data which contains more information. On the one hand, using only year-end data of ETF ownership will ignore ETF activity in non-reporting periods. On the other hand, using daily data instead of intraday data to calculate pricing efficiency will lose intraday trading information, especially noise information, leading to overestimation of pricing efficiency.

Second, we change the dependent variables by using a new pricing error proxy *PE4* and a new variance ratio *VR(30)*. As shown in Panel A of Table 6, ETF ownership still significantly reduces the pricing efficiency of the underlying stocks.

Third, we exploit the first-order difference of the ETF ownership $\Delta ETF\_ratio_{\ i,t}$ as the independent variable. The results in Panel B of Table 6 show that the increase in ETF ownership is accompanied by a decrease in the pricing efficiency of the underlying stocks, which is consistent with our benchmark results.

Fourth, we eliminate the samples in June and December due to the adjustment of index constituents. The results in Panel C of Table 6 show the negative impact of ETF ownership on the pricing efficiency of constituent stocks. Overall, our results are robust to alternative proxies and other subsamples.

5. Conclusion

As a rapidly growing financial innovation product, ETFs enrich the investment tools of the underlying stocks but also bring non-fundamental shocks to them. Utilizing the intraday trading data of stocks and daily ETF data in the Chinese market from 2012 to 2021, we find that the increase in ETF ownership reduces the pricing efficiency of the underlying stocks. Channel test show that ETF ownership enhances the intensity of ETF arbitrage activities, which exacerbate the contagion of noise information from the ETF market

Table 6
Other robustness test.

Panel A: Alternative dependent variable	PE4			VR(30)		
	(1)	(2)	(3)	(4)	(5)	(6)
ETF_ratio (t-1)	0.0053***	0.0035***	0.0030***	0.0015***	0.0011***	0.0014***
	(12.20)	(8.90)	(9.77)	(5.26)	(3.42)	(3.74)
Lag(y)			0.2294***			0.0038***
			(50.64)			(7.59)
Control	YES	YES	YES	YES	YES	YES
Firm fixed effect	NO	YES	YES	NO	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES
Observations	5,136,608	5,136,605	3,890,334	5,384,844	5,384,841	4,233,030
Adjusted R2	0.322	0.364	0.407	0.017	0.018	0.018
Panel B: Alternative independent variable	PE2			VR(5)		
	(1)	(2)	(3)	(4)	(5)	(6)
ΔETF_ratio	0.0080***	0.0076***	0.0052***	0.0923***	0.0882***	0.0341***
	(11.68)	(11.31)	(7.03)	(8.22)	(7.96)	(2.83)
Lag(y)			0.2676***			0.1851***
			(66.01)			(69.58)
Control	YES	YES	YES	YES	YES	YES
Firm fixed effect	NO	YES	YES	NO	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES
Observations	5,136,608	5,136,605	3,890,334	5,385,496	5,385,493	4,233,752
Adjusted R2	0.381	0.429	0.486	0.350	0.378	0.404
Panel C: Alternative subsamples	PE2			VR(5)		
	(1)	(2)	(3)	(4)	(5)	(6)
ETF_ratio (t-1)	0.0051***	0.0029***	0.0024***	0.0619***	0.0416***	0.0342***
	(13.22)	(8.58)	(9.08)	(13.20)	(9.83)	(9.70)
Lag(y)			0.2664***			0.1838***
			(62.05)			(68.21)
Control	YES	YES	YES	YES	YES	YES
Firm fixed effect	NO	YES	YES	NO	YES	YES
Time fixed effect	YES	YES	YES	YES	YES	YES
Observations	4,233,538	4,233,536	3,204,888	4,435,996	4,435,994	3,480,496
Adjusted R2	0.392	0.439	0.487	0.343	0.373	0.401

This table reports the results of robust regression results including replacing the dependent variable, replacing independent variable and eliminating the samples in June and December. We change the dependent variables by using a new pricing error proxy PE4 and a new variance ratio VR(30) in Panel A. And we exploit the first-order difference of the ETF ownership as the independent variable in Panel B. Finally, we eliminate the samples in June and December due to the adjustment of index constituents in Panel C. Values in parentheses are the *t*-values.***, **, and * denote significance at the 1%, 5%, and 10% levels, respectively.

to the stock market, reducing the pricing efficiency of the underlying stocks. To address the endogeneity issue, we adopt the ETF policy change from the Shenzhen Stock Exchange in 2019 as an exogenous shock to ETF arbitrage and conduct the instrumental variable (IV) method to corroborate our main findings. In addition, we compare our results with prior empirical work and attribute the differences to the more granular data and method we use. Overall, our study provides a new method to estimate ETF holdings and adds to the literature on the negative consequences of ETFs.

However, there are some aspects of improvement in this paper. On the one hand, we only consider the impact of non-cross-border and passive ETF ownership on constituents, ignoring the rapid development of cross-border ETFs and active ETFs in the Chinese market from 2022. On the other hand, arbitrage activities are transient and our conclusions could be improved by using higher frequency data in the identification strategy.

CRediT authorship contribution statement

Guanhua Chen: Conceptualization, Writing – review & editing. **Xiangli Liu:** Supervision, Methodology, Formal analysis, Conceptualization. **Xiao Liu:** Writing – original draft, Writing – review & editing. **Zhihua Zhao:** Writing – original draft, Software, Project administration, Investigation, Data curation, Writing – review & editing.

Data availability

Data will be made available on request.

References

Aldridge, I., 2016. ETFs, high-frequency trading, and flash crashes. J. Portfolio Manag. 43 (1), 17–28.
Antoniou, C., Weikai Li, F., Liu, X., Subrahmanyam, A., Sun, C., 2023. Exchange-traded funds and real investment. Rev. Financ. Stud. 36 (3), 1043–1093.
Appel, I.R., Gormley, T.A., Keim, D.B., 2016. Passive investors, not passive owners. J. Financ. Econ. 121 (1), 111–141.
Ben-David, I., Franzoni, F., Moussawi, R., 2018. Do ETFs increase volatility? J. Finance 73 (6), 2471–2535.
Ben-David, I., Franzoni, F., Moussawi, R., 2017. Exchange-traded funds. Annu. Rev. Financ. Econ. 9, 169–189.
Boone, A.L., White, J.T., 2015. The effect of institutional ownership on firm transparency and information production. J. Financ. Econ. 117 (3), 508–533.
Box, T., Davis, R., Evans, R., Lynch, A., 2021. Intraday arbitrage between ETFs and their underlying portfolios. J. Financ. Econ. 141 (3), 1078–1095.
Brogaard, J., Nguyen, T.H., Putnins, T.J., Wu, E., 2022. What moves stock prices? The roles of news, noise, and information. Rev. Financ. Stud. 35 (9), 4341–4386.
Brown, D.C., Davies, S.W., Ringgenberg, M.C., 2021. ETF arbitrage, non-fundamental demand, and return predictability. Rev. Financ. 25 (4), 937–972.
Da, Z., Shive, S., 2018. Exchange traded funds and asset return correlations. Eur. Financ. Manag. 24 (1), 136–168.
Dannhauser, C.D., 2017. The impact of innovation: evidence from corporate bond exchange-traded funds (ETFs). J. Financ. Econ. 125 (3), 537–560.
Glosten, L., Nallareddy, S., Zou, Y., 2021. ETF activity and informational efficiency of underlying securities. Manag. Sci. 67 (1), 22–47.
Goldstein, I., 2023. Information in financial markets and its real effects. Rev. Financ. 27 (1), 1–32.
Hasbrouck, J., 1993. Assessing the quality of a security market: a new approach to transaction-cost measurement. Rev. Financ. Stud. 6 (1), 191–212.
Huang, S., O'Hara, M., Zhong, Z., 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. Account. Stud. 22 (3), 1048–1083.
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.
Marshall, B.R., Nguyen, N.H., Visaltanachoti, N., 2013. ETF arbitrage: intraday evidence. J. Bank. Financ. 37 (9), 3486–3498.
Pan, K., Zeng, Y., 2017. ETF arbitrage under liquidity mismatch. Available at SSRN 3723406.
Rösch, D.M., Subrahmanyam, A., Van Dijk, M.A., 2017. The dynamics of market efficiency. Rev. Financ. Stud. 30 (4), 1151–1187.
Wu, W., Zhu, F., 2023. ETF ownership and informational efficiency of underlying stocks: evidence from China. Pacific-Basin Finance J. 79, 102005.
Xiong, X., Gao, Y., Feng, X., 2017. Successive short-selling ban lifts and gradual price efficiency: evidence from China. Account. Finance 57 (5), 1557–1604.