ACCOUNTING
& FINANCE

Accounting & Finance 58 (2019) 1525–1548

Do exchange-traded fund flows increase the volatility of the underlying index? Evidence from the emerging market in China

Hua Wang[a], Liao Xu[b]

[a]*School of Accounting, Zhongnan University of Economics and Law, Wuhan, China*
[b]*Business School, East China University of Political Science and Law, Shanghai, China*

Abstract

Studying 70 Chinese equity exchange-traded funds (ETFs), we show that daily ETF flows significantly increase both the total volatility and the fundamental volatility of the underlying index on the next trading day. More specifically, it is the forward-looking flow component which captures APs' share creation/redemption activities beyond their role of market makers that can significantly predict the two types of volatility. Moreover, ETF arbitrage (ETF's information share) enhances the effect of forward-looking flows on the total volatility (fundamental volatility) of the index. Furthermore, the relationships between forward-looking flows and the two types of index volatility show a two-way contagion.

Key words: Exchange-traded fund; Fund flows; Trading volume; Stock market volatility; Efficient price

JEL classification: G12, G14

doi: 10.1111/acfi.12437

1. Introduction

Flows of exchange-traded funds (ETFs) are the result of ETF share creation/redemption, which is known as an 'in kind' exchange initiated by authorised participants (APs) in the primary market. New ETF shares are created by depositing a portfolio of stocks that closely approximates the ETF holdings, and outstanding shares can be redeemed in exchange for the basket portfolio. As index funds, ETFs replicate the performance of the designated indexes as closely as possible. Issuers and exchanges set forth diversification opportunities to all types of investors at a lower cost, with more efficient taxation and higher

Please address correspondence to Liao Xu via email: L3xu@outlook.com

transparency, contrary to conventional mutual funds. The 'in-kind' creation/redemption mechanism is not only the way ETFs gain exposure to the market but also the 'secret source' supporting these advantages.[1] Therefore, ETF flows have attracted increasing attention in recent years. In this study, we exploit the emerging ETF market in China to examine whether and how ETF flows affect the volatility of the ETF's index by looking into specific components of ETF flows.

We focus on index volatility in terms of both total volatility and fundamental volatility. While it is well recognised that trading affects price/return changes, less is known about whether and how ETF flows, the consequence of APs' trading activities in both primary and secondary markets, affect the volatility of the underlying index. Total volatility is the observed change of price/return over a period regardless of the causes of the change, while fundamental volatility is driven by the information including both private and public information and is not observable. Thus, differently motivated ETF flows may affect the two types of volatility in different ways. The volatility of the underlying index is the focus of this paper because the 'in-kind' exchange warrants a cross-market shock contagion across ETFs and their indexes. Volatility is also an important topic on its own because it affects return autocorrelation (McKenzie and Faff, 2003), liquidity (Chordia *et al.*, 2005) and momentum (Wang and Xu, 2012), among others.

We look at the Chinese ETF market for two reasons. First, China has been the second largest economy since 2010 and is now playing the role of the unambiguously largest emerging market worldwide. The growth of Chinese local ETFs is dramatic. Since the debut of the first Chinese ETF, Huaxia SSE 50 ETF (ticker: 510050), in February 2005, the total assets under management (AUMs) of Chinese ETFs have exceeded 232 billion yuan and the total number of ETFs reached more than 140 by the end of 2017, according to Wind.[2] Such a rapidly growing market arouses great interest among not only academics but also practitioners and regulators, all of which motivates us in this study. Second, the trading rule of $T + 1$ implemented in the Chinese stock market facilitates the specification of different motives for APs' share creation/redemption activities, although this rule reduces efficiency (Chen *et al.*, 2017),[3] as our study addresses the distinct roles of APs. ETF shares are usually created/redeemed at closing or after hours.

[1] See, for instance, Madhavan *et al.* (2014), Lettau and Madhavan (2016), and Madhavan (2016).

[2] According to Wind, equity ETFs account for more than 90% of all types of ETFs in China.

[3] Since the mid-1990s, the Chinese stock market has been subject to the $T + 1$ rule, under which investors must hold the stocks for at least one trading day before they can sell it.

To the best of our knowledge, this paper is not only the first that explicitly relates ETF flows and their specific decomposed components to the different types of volatility of the underlying index, but it is also a pioneer in comprehensively studying the emerging ETF market in China regarding this relationship. More specifically, we decompose an ETF's daily flows by regressing them against the concurrent market demand for the ETF. This decomposition process estimates a flow component due to the APs' duty of market-making in response to demand, leaving the residuals to be presumably driven by other motives beyond the response to demand. The first component is backward-looking as it passively reflects the ETF demand in the secondary market, while the second component is forward-looking as it captures APs' share creation/redemption motives beyond market-making on their own accounts. Then, we examine their particular volatility effects. The major contributions of this study are summarised as follows. First, we confirm a predictive relationship between ETF flows and the total volatility of the ETF's index in the sense that ETF flows significantly increase the index's total volatility on the next trading day. More specifically, the backward-looking flow component only exerts a slight predictive impact on the index's total volatility, whereas such an impact by the forward-looking flow component is highly profound.

Second, the paper sheds light onto index fluctuation driven by shocks to the unobserved fundamentals. We quantify the fundamental volatility by the variance of permanent component of price shocks to the index through the Beveridge–Nelson (hereafter BN) decomposition (Beveridge and Nelson, 1981). For robustness, we also follow Hasbrouck's (1995) framework to estimate the fundamental fluctuation by the variance in efficient price innovations. Empirical evidence shows that ETF flows can also increase the fundamental volatility of an index on the next trading day. Moreover, it is the forward-looking component of ETF flows that largely accounts for this predictive relationship, whereas the role played by the backward-looking component is rather inclusive in predicting fundamental volatility. This implies that the forward-looking flows may contain new information about fundamentals and this information is then revealed to the index market on the next trading day.

Third, considering arbitrage and information as APs' potential motives for forward-looking flows of ETFs, we also include ETF mispricing and the ETF's information share in the tests. The results explain that ETF arbitrage stimulates the predictive power of forward-looking flows on the index's total volatility but not on the fundamental volatility. Conversely, an ETF's information share enhances the effect of forward-looking flows on the index's fundamental volatility but not on the total volatility. This therefore suggests that the APs' share creation/redemption activities, motivated by different driving forces, affect the two types of index volatility in different ways: the force of arbitrage accounts for the observed change in the index while the force of information exerts the effect on the unobserved fluctuation driven by shocks to the fundamental value. On the other hand, neither ETF arbitrage nor ETF's information share exerts a significant impact on the path through which the backward-looking flows affect volatility. This is because the backward-looking flows solely rely on investor demand and are orthogonal to APs' other motives such as arbitrage and information.

Fourth, we also provide evidence that a two-way Granger causality exists between ETF flows (the forward-looking flow component in particular) and the two types of index volatility. More specifically, we show the forward-looking flows Granger-cause the ETF's index to be more (fundamentally) volatile and, on the other hand, a more (fundamentally) volatile index also Granger-causes more forward-looking ETF flows. It thus reinforces the evidence supporting a close correlation and strong interaction between forward-looking flows and the two types of index volatility.

The rest of the paper proceeds as follows. Section 2 provides an overview of the related literature. It also discusses the motivation and intuition behind the empirical analysis. The ETF sample, flow decomposition, and regression variables are described in Section 3. Sections 4 and 5 examine the index's total volatility and fundamental volatility, respectively. Granger causality is detailed in Section 6. Section 7 summarises the robustness checks, and the last section concludes the paper.

2. Related literature, motivation and intuition

2.1. Related literature

The great success of ETFs attracts the attention of academics in relation to not only ETF markets themselves but also their impacts on securities in ETF baskets. For example, studying the relationship between ETFs and the volatility of the underlying securities, Stratmann and Welborn (2012) study the factors leading to ETF fails-to-deliver and they argue that this failure Granger-causes higher index volatility. Krause *et al.* (2014) demonstrate that volatility flows from ETFs to their largest underlying stocks. Xu and Yin (2017b) reveal the relationship between ETF trading volume and the volatility of the underlying index. Ben-David *et al.* (2017) explore a positive correlation between the volatility of a stock and its ownership by ETFs.

Focusing on ETF flows, Clifford *et al.* (2014) assert that ETF flows are generally driven by investor demand. Staer (2017) associates ETF flows with the returns on their indexes. Ben-David *et al.* (2017) and Brown *et al.* (2018) attribute ETF flows to arbitrage activity of APs. Broman and Shum (2018) find that ETF flows can be predicted by ETF liquidity relative to the liquidity of the underlying basket. Xu *et al.* (2018) emphasise the information roles of APs and analyse and confirm the predictability of information-driven ETF flows on ETF returns. Nevertheless, there are few explicit explorations of the relationship between ETF flows, particularly the specific components of ETF flows, and the volatility of the underlying index in the literature.

Looking at the emerging ETF market in China, Hua *et al.* (2016) study the Jiashi CSI 300 ETF and CSI 300 index futures, and demonstrate that the pre-open and post-close futures trades influence ETF returns. Chen *et al.* (2017) investigate the effect of the trading rule of $T + 1$ for the Chinese stock market on speculation by studying the Huatai and Jiashi CSI 300 ETFs. While these two CSI 300 ETFs are large ETFs in China, they are such singular trading instruments that they may not be generalisable to a broader market. This paper contributes to the local conversation not only by exploring the correlation between ETF flows and index volatility but also in a comprehensive fashion in terms of an extensive inclusion of most of the Chinese equity ETFs.

2.2. Motivation and intuition

Most equity ETFs are investment vehicles passively tracking a benchmark index. They are listed on a stock exchange and traded on the open market. ETF shares are created by APs by acquiring all underlying constituents in the same weights as the index and exchanging with the ETF issuer for the ETF shares with the equivalent value. Likewise, an AP redeems ETF shares by delivering the funds to the ETF issuer in exchange for the equivalent value of constituent stocks in proportion to the index's weights. The creation/redemption process actually represents a mechanism of shock contagion between financial markets (Kyle and Xiong, 2001; Greenwood, 2005).

Exchange-traded fund flows are the result of ETF share creation/redemption activities undertaken by APs. In the existing literature, there are three major motives for APs to create/redeem ETF shares. The first is the ETF demand in the secondary market. Part of the usual role of APs is to serve as market makers, and they can create/redeem shares for rebalancing the positions in response to the market demand on the day. As Clifford *et al.* (2014) indicate, 'ETF flow should be interpreted generally as investor demand, as share creation implies the purchase of shares by somebody.' Second, APs' trading activities for their own arbitrage profits also account for ETF flows; see Ben-David *et al.* (2017) and Brown *et al.* (2018).[4] Xu *et al.* (2018) demonstrate that APs are informed traders and can create/redeem ETF shares in anticipation of upcoming news. The first motive is backward-looking, but the latter two are forward-looking. Therefore, this paper decomposes ETF flows into backward-looking and forward-looking components. The former features APs' role as market makers in response to the investor demand, while the latter captures

[4] APs can arbitrage any sizeable differences between the ETF price and its net asset value (NAV). For example, with the ETF price below its NAV, APs buy ETFs in the secondary markets and take on a short position in the underlying index constituents and then redeem the ETFs for the basket constituents in the primary market before closing the short position at an arbitrage profit.

APs' creation/redemption activities on their own accounts, beyond their role as market makers.[5]

Intuitively, the backward-looking component of ETF flows is passive and does not contain APs' beliefs about the markets in the future. Such beliefs are, however, contained in the forward-looking component and likely to be propagated to the secondary markets through subsequent trading activities (e.g., Xu and Yin, 2017b). In this sense, the forward-looking component that contains both arbitrage-driven and information-driven flows should exert the predictive power on both the total volatility and fundamental volatility of the ETF's index. However, the predictive power of the backward-looking component on either type of volatility should not be considerable.

3. Sample ETFs, flow decomposition, and regression variables

3.1. Sample ETFs

Our sample includes a broad selection of Chinese passively managed equity ETFs from the Wind database from January 2015 to December 2017. The three-year-period is due to the availability of high-frequency data.[6] When selecting ETF samples, we required them to have at least 3.5 years of trading history; that is, only those ETFs born on or before July 2014 would be selected. This is because the number of ETF shares outstanding in their early life cycle can be unreliable because newly created ETFs experience dramatic creation/redemption activities as indicated by Broman and Shum (2018). After screening, our final sample includes 70 distinct ETFs.

3.2. Flow decomposition

Since ETF shares are created/redeemed at the close of the day or after hours, $Flow_{i,t}$ is defined by the absolute value of the difference between ETF shares outstanding on day t and $t+1$:

$$Flow_{i,t} = \left| \frac{Shares\ Outstanding_{i,t+1} - Shares\ Outstanding_{i,t}}{Shares\ Outstanding_{i,t}} \right|.$$

The data of shares outstanding are obtained from the Wind economic database. The descriptive statistics of $Flow_{i,t}$ are shown in Panel B of Table 1.

[5] These three motives are best supported by the existing literature. While there might exist other motives for APs to create/redeem ETFs besides these three, their roles are vague (e.g., Xu *et al.*, 2018). Identifying further motives is, however, an open empirical question for further research.

[6] Wind and other database in China only provide the high-frequency data within the nearest three years.

To estimate the ETF flow components, we run a time-series regression of $Flow_{i,t}$ against the concurrent market demand of ETFs proxied by ETF turnover, $Turnover_{i,t}$, for each of the 70 ETFs:

$$Flow_{i,t} = a + \beta_i Turnover_{i,t} + e_{i,t}, \tag{1}$$

where

$$Turnover_{i,t} = \frac{Trading\ Volume_{i,t}}{Shares\ Outstanding_{i,t}}$$

The averaged results of (1) across 70 ETFs are reported in Table 2, in which the average coefficients, t-statistics, and adjusted R^2 are displayed. Regression coefficients of $Turnover_{i,t}$ are significantly positive at least at the 10 percent level for 100 percent of our sample ETFs. They are on average significant at the 1 percent level with the average adjusted R^2 of the regression models reaching 39 percent. This indicates that 39 percent of the ETF flows can be explained by the concurrent market demand.

After estimating regression model (1) for each ETF, ETF flows driven by market demand is thus defined by $Flow_{i,t}^{backward} = \hat{\beta}_i Turnover_{i,t}$ where $\hat{\beta}_i$ is the estimate of β_i in (1). We call this the backward-looking component of ETF flows. Moreover, the ETF flows driven by APs' other motives beyond $Flow_{i,t}^{backward}$ is estimated through the residuals of (1), i.e., $Flow_{i,t}^{forward} = \left|\hat{e}_{i,t}\right|$, which is orthogonal to APs' roles as market makers. We call this the forward-looking flow component.

Since one may be concerned that APs are also likely to respond to the market demand of previous days, we include the first and second lags of $Turnover_{i,t}$ into the regressions. The extended exercises also include the first and second lags of $Flow_{i,t}$ for its potential autocorrelation. The results for these extended exercises are reported in the second to fourth columns of Table 2. Apparently, none of the effects of these lagged terms is significant, and, in turn, we focus on regression (1) for decomposition.

3.3. Regression variables

We use realised variance of index intraday returns for the daily measure of index's total volatility. We compute the realised volatility based on 1-min and 5-min index returns.[7] Estimating the fundamental volatility, we conduct the BN decomposition (Beveridge and Nelson, 1981) and Hasbrouck's (1995) framework. The estimation details are given in Section 5. The descriptive statistics for the volatility measures are given in Panel A of Table 1.

[7] Realised volatility data are scaled up by 10,000 after computation.

Table 1
Descriptive statistics

Panel A: Descriptive statistics of dependent variables

	$RV_{i,t}^{1min}$	$RV_{i,t}^{5min}$	$FV_{i,t}^{BN}$	$FV_{i,t}^{HAS}$
Mean	0.375	0.468	0.102	0.124
Median	0.219	0.282	0.093	0.103
Max	2.541	2.732	0.757	0.936
Min	0.020	0.018	0.005	0.007
Std. dev	0.427	0.512	0.131	0.140

Panel B: Descriptive statistics of ETF flows and decomposed flow components

	$Flow_{i,t}$	$Flow_{i,t}^{backward}$	$Flow_{i,t}^{forward}$
Mean	0.012	0.011	0.008
Median	0.005	0.006	0.004
Max	0.286	0.159	0.225
Min	0.000	0.001	0.000
Std. dev	0.024	0.015	0.021

Panel C: Descriptive statistics of control variables

	$Illq_{i,t}$	$Volume_{i,t}$	$Mk_{i,t}$
Mean	0.009	0.021	0.031
Median	0.006	0.014	0.029
Max	0.238	0.305	0.060
Min	0.000	0.003	0.022
Std. dev	0.013	0.030	0.046

Panel D: Descriptive statistics of ETF arbitrage and ETF's information share

	$Arb_{i,t}$	$IS_{i,t}$
Mean	0.921	78.659
Median	0.720	89.569
Max	10.429	99.244
Min	0.003	0.528
Std. dev	0.938	26.021

The table reports the descriptive statistics of the key variables used in our analysis. Panel A is of the dependent variables, including the total volatility measures and fundamental volatility measures of the index. $RV_{i,t}^{1\,\min}$ and $RV_{i,t}^{5\,\min}$ are the realised variance of 1-min and 5-min index returns, respectively. $FV_{i,t}^{BN}$ and $FV_{i,t}^{HAS}$ are the fundamental volatility measures estimated by BN decomposition and Hasbrouck's (1995) model, respectively. Panel B is of the daily ETF flow $Flow_{i,t}$ and its backward-looking component $Flow_{i,t}^{backward}$ and forward-looking component $Flow_{i,t}^{forward}$. Panel C displays the statistics for control variables, Amihud's (2002) illiquidity of the index $Illq_{i,t}$, index trading volume $Volume_{i,t}$, and the market capitalisation of the index $Mk_{i,t}$. Panel D exhibits the statistics of ETF arbitrage $Arb_{i,t}$ and ETF's information share $IS_{i,t}$. The sample period of an ETF starts in January 2015 and ends in December 2017. All variables are first averaged over their time-series and then their cross-sectional descriptive statistics are reported in the table.

Table 2
Exchange-traded fund flow decomposition

	$Flow_{i,t}$			
	(1)		Extensions on (1)	
Constant	0.0002	0.000	0.000	0.0001
	(0.27)	(0.01)	(0.02)	(0.16)
$Turnover_{i,t}$	0.235***	0.210***	0.209***	0.206***
	(21.56)	(18.08)	(17.29)	(16.20)
	100.00%+	100%+	100.00%+	100.00%+
	0.00%−	0.00%−	0.00%−	0.00%−
$Turnover_{i,t-1}$		0.033	0.032	0.038
		(1.16)	(1.01)	(1.24)
		22.86%+	21.43%+	22.86%+
		5.71%−	4.29%+	4.29%−
$Turnover_{i,t-2}$			0.018	−0.009
			(0.54)	(−0.26)
			11.43%+	8.57%+
			10.00%−	7.14%−
$Flow_{i,t-1}$				0.066
				(1.49)
				25.71%+
				4.29%−
$Flow_{i,t-2}$				0.042
				(0.96)
				14.29%+
				7.14%−
Adjusted R^2	0.39	0.39	0.39	0.39

The table reports the results of the time-series regressions of ETF daily flow, $Flow_{i,t}$, for each of the 70 sample ETFs. Explanatory variable $Turnover_{i,t}$ is the ETF turnover. The sample period of an ETF starts in January 2015 and ends in December 2017. The table reports the estimated coefficients averaged across all ETFs. Average t-statistics are in parentheses. *** indicate the overall significance at the 1, 5 and 10 percent levels, respectively. Then, it reports the percentages of positive and negative coefficients that are significant at least at the 10 percent level and are indicated respectively by %+ and %−.

Control variables in the analysis include Amihud's (2002) illiquidity ratio of the index, and index capitalisation.[8] Inspired by the well-known volume–volatility relationship in single stock markets, we further include the index trading volume as a control.[9] To capture the volatility persistence, the first lag

[8] See Chordia *et al.* (2001, 2005), and Weber and Rosenow (2006).

[9] The study of the volume–volatility relationship has a long history, dating back to Granger and Morgenstern (1963) and Clark (1973). Karpoff (1987) provides a comprehensive review of the early literature exploring the volume–volatility relationship.

of a volatility measure is also included.[10] The data of the control variables are also obtained from the Wind database. Because the time trends of all variables over the 2015–2017 period are not severe, we keep their raw series in the analysis.[11] Panel C of Table 1 summarises the descriptive statistics for the control variables.

4. ETF flows and index's total volatility

In this section, we examine the relationship between ETF flows and the total volatility of the ETF's index, and then we look into the particular volatility effects of the two flow components. To be specific, we estimate the following panel regressions with ETF fixed effects and month fixed effects included:

$$Y_{i,t+1} = \alpha + \beta Flow_{i,t} + \gamma^j Controls^j_{i,t+1} + \pi Y_{i,t} + e_{i,t}, \tag{2}$$

$$Y_{i,t+1} = \alpha + \beta_1 Flow^{backward}_{i,t} + \beta_2 Flow^{forward}_{i,t} + \gamma^j Controls^j_{i,t+1} + \pi Y_{i,t} + e_{i,t}, \tag{3}$$

where $Y_{i,t+1}$ has two specifications: the realised variance of 1-min returns of index i on day $t + 1$, $RV^{1\,\min}_{i,t+1}$, and its counterpart computed using 5-min returns, $RV^{5\,\min}_{i,t+1}$. In (2), $Flow_{i,t}$ is undecomposed end-of-day flows of ETF i on day t. In (3), $Flow^{backward}_{i,t}$ and $Flow^{forward}_{i,t}$ are the backward-looking and forward-looking flow components, respectively. Control variables $Controls^j_{i,t+1}$ in turn include $Illq_{i,t+1}$, $Volume_{i,t+1}$, and $Mk_{i,t+1}$. They are, respectively, Amihud's (2002) illiquidity ratio, trading volume, and capitalisation of index i on day $t + 1$. To capture the persistence of volatility, we include Y_t in the regressors. The results of (2) and (3) are respectively reported in Panels A and B of Table 3.[12]

From the first panel, it is apparent that ETF flows increase index's total volatility on the next trading day. While one may attribute this observation to the high correlation between the trading activities in index markets and the ETF share creation/redemption mechanism, we show that the coefficients of ETF flows are highly significant after controlling for the index trading volume in the regressors. It suggests that ETF flows have their own impact on index volatility independent of index trading volume. To further examine this flow–volatility relationship, we conduct regression (3) using decomposed flow components. In

[10] For volatility persistence, see Chou (1988) and Lamoureux and Lastrapes (1990), who empirically show the persistence of shocks to volatility by GARCH models.

[11] All variables survive the augmented Dickey–Fuller (ADF) unit root test for stationarity.

[12] In addition to panel analysis, we also conducted the time-series analysis for each of the 70 ETFs. There is no qualitative change in the major results between the two analyses. To restrict the number of tables, the time-series analysis is not reported in the paper but available upon request from the authors.

Table 3
The predictive relationship between (decomposed) ETF flows and index's total volatility

Panel A: The effects of undecomposed ETF flows

	$RV^{1min}_{i,t+1}$	$RV^{5min}_{i,t+1}$
$Flow_{i,t}$	3.625***	4.352***
	(4.75)	(4.86)
$Illq_{i,t+1}$	4.542***	4.925***
	(3.34)	(3.22)
$Volume_{i,t+1}$	3.388***	3.253***
	(13.21)	(11.25)
$Mk_{i,t+1}$	−7.334***	−7.724***
	(−3.56)	(−3.28)
Constant	−0.112	−0.205
	(0.68)	(0.74)
First lag of dependent variable	0.197***	0.188***
	(33.35)	(32.27)
ETF fixed effect	Yes	Yes
Month fixed effect	Yes	Yes
Adj. R^2	0.26	0.25

Panel B: The effects of decomposed ETF flows

	$RV^{1min}_{i,t+1}$	$RV^{5min}_{i,t+1}$
$Flow^{backward}_{i,t}$	1.378	1.424
	(1.57)	(1.36)
$Flow^{forward}_{i,t}$	3.592***	4.203***
	(5.05)	(5.27)
$Illq_{i,t+1}$	4.426***	4.830***
	(3.21)	(3.14)
$Volume_{i,t+1}$	3.240***	3.039***
	(12.64)	(10.52)
$Mk_{i,t+1}$	−7.036***	−7.452***
	(−3.32)	(−2.85)
Constant	−0.125	−0.224
	(−0.79)	(0.90)
First lag of dependent variable	0.190***	0.185***
	(32.16)	(31.65)
ETF fixed effect	Yes	Yes
Month fixed effect	Yes	Yes
Adj. R^2	0.26	0.25

Panel A displays the results of the panel regressions of index's total volatility on day $t+1$ against the undecomposed ETF flows, $Flow_{i,t}$, on day t. Dependent variables $RV^{1min}_{i,t+1}$ and $RV^{5min}_{i,t+1}$ are the realised variance of 1-min and 5-min index returns, respectively, on day $t+1$. Control variables include Amihud's (2002) illiquidity of the index $Illq_{i,t}$, index trading volume $Volume_{i,t}$, and the market capitalisation of the index $Mk_{i,t}$. ETF fixed effect and month fixed effect apply. In Panel B, ETF flows, $Flow_{i,t}$, is replaced by its two decomposed components: backward-looking component $Flow^{backward}_{i,t}$ and forward-looking component $Flow^{forward}_{i,t}$, other settings being the same as those in Panel A. The sample period of an ETF starts in January 2015 and ends in December 2017. The t-statistics are reported in parentheses. *** indicate significance at the 1, 5 and 10 percent levels, respectively.

Panel B, we find that the backward-looking flow component only exerts limited predictive power in the index's total volatility, whereas the corresponding power of the forward-looking component is highly significant. It shows that $RV^{1\,\min}_{i,t+1}$ increases by 18 percent of its standard deviation for an increment in $Flow^{forward}_{i,t}$ by one standard deviation, whereas the corresponding increase is only 5 percent for $Flow^{backward}_{i,t}$.[13] This suggests that APs' response to market demand adds little power in predicting the index's total volatility beyond the volatility effect of the index trading volume, which is controlled for in the regressors.[14] However, even with the presence of index trading volume, the forward-looking component that features APs' creation/redemption motives beyond their role as market makers can still predict the index's total volatility significantly.

For control variables, the results indicate that smaller indexes are associated with greater volatility. Positive and significant coefficients of index trading volume confirm the widely known volume–volatility relationship. The estimates on illiquidity suggest that illiquidity has a positive correlation with index volatility. Lagged Y_t is highly significant and positive for all settings, which suggests the strong persistence of index volatility. These findings about controls are all consistent with the existing literature (see footnotes 4–6).

The forward-looking component of ETF flows is assumed to be driven by APs' trading motives on their own accounts in addition to their role as market makers. Two of such motives can be arbitrage (see Petajisto, 2016; Ben-David *et al.*, 2017; Brown *et al.*, 2018) and information (see Xu *et al.*, 2018). Following the insight shed by Ben-David *et al.* (2017) that end-of-day ETF mispricing is a reliable indicator of the next day's arbitrage, we include the ETF mispricing at the close of a day, $Arb_{i,t}$ to the regressors. $Arb_{i,t}$ is defined by:

$$Arb_{i,t} = \left| \frac{Close_{i,t} - NAV_{i,t}}{Close_{i,t}} \right| \times 1000,$$

where $Close_{i,t}$ and $NAV_{i,t}$ are, respectively, the closing price and the net asset value of ETF i at the close of day t. In addition, Xu *et al.* (2018) argue that the information share of an ETF increases with its information-driven flows, and therefore, we also include the ETF's information share IS from Hasbrouck (1995) to the regressors. Daily $IS_{i,t}$ is estimated through the bivariate vector error correction model (VECM) between an ETF and its index at a 1-min frequency over 30 lags, and then scaled up by 100.[15] The descriptive statistics of

[13] $3.592 \times \frac{0.021}{0.427} = 0.177; 1.378 \times \frac{0.015}{0.427} = 0.048.$

[14] In the unreported exercises, the backward-looking component becomes significant while the trading volume of the index is excluded in the regressors. It implies that the backward-looking component somehow acts as a proxy for index trading activities when the trading volume of the index is absent in the model.

[15] Following Baillie *et al.* (2002), we adopt the average of Hasbrouck's (1995) information share upper and lower bounds as *IS*.

$Arb_{i,t}$ and $IS_{i,t}$ are reported in Panel D of Table 1. To more clearly explore the capability of the driving forces of inter-market arbitrage and information in affecting the relationship between the two ETF flow components and the index volatility, we employ their interaction terms with the two flow components.

Thus, together with $Arb_{i,t}$ and $IS_{i,t}$, the four interaction terms, $Flow_{i,t}^{backward} \times Arb_{i,t}$, $Flow_{i,t}^{forward} \times Arb_{i,t}$, $Flow_{i,t}^{backward} \times IS_{i,t}$, and $Flow_{i,t}^{forward} \times IS_{i,t}$, are also included in the regressors:

$$\begin{aligned} Y_{i,t+1} = \alpha &+ \beta_1 Flow_{i,t}^{backward} + \beta_2 Flow_{i,t}^{forward} + \beta_3 Arb_{i,t} + \beta_4 Flow_{i,t}^{backward} \times Arb_{i,t} \\ &+ \beta_5 Flow_{i,t}^{forward} \times Arb_{i,t} + \beta_6 IS_{i,t} + \beta_7 Flow_{i,t}^{backward} \times IS_{i,t} \\ &+ \beta_8 Flow_{i,t}^{forward} \times IS_{i,t} + \gamma^j Controls_{i,t+1}^j + \pi Y_{i,t} + e_{i,t} \end{aligned} . \tag{4}$$

The results of the extended regression (4) are reported in Table 4.

From the table, we show that with a greater end-of-day arbitrage opportunity, the forward-looking component of ETF flows exerts a magnified impact on the index's total volatility on the next trading day, evidenced by positive and significant coefficients of $Flow_{i,t}^{forward} \times Arb_{i,t}$ in both model specifications. It points to the fact that a greater arbitrage opportunity stimulates APs to create/redeem more ETF shares for their own arbitrage profits and thus intensify the effect of ETF flows on volatility prediction. However, insignificant coefficients of $Flow_{i,t}^{forward} \times IS_{i,t}$ suggest that an ETF's information share cannot substantially increase the predictive power of $Flow_{i,t}^{forward}$ on the index's total volatility. On the other hand, share creation/redemption for market-making purposes is only in response to investor demand and orthogonal to APs' trading motives on other accounts, such as arbitrage profit or information advantages. Therefore, it is very intuitive that the coefficients of $Flow_{i,t}^{backward} \times Arb_{i,t}$ and $Flow_{i,t}^{backward} \times IS_{i,t}$ are insignificant here, as shown in Table 4.

In a nutshell, we document that ETF flows can significantly predict the index's total volatility. Such predictive power is largely driven by APs' share creation/redemption activities for their own advantages rather than by their passive response to market demand. Moreover, when ETF mispricing becomes more sizable at the end of a day, the predictive power of the forward-looking component on the index's total volatility on the next trading day is magnified. However, no evidence is found that the backward-looking flow component yields a significant impact on the index's total volatility after controlling for the index trading volume.

5. ETF flows and fundamental volatility

The total volatility only gauges the observed price/return variation but cannot explore the fluctuation of the fundamental value which is hidden in the observed price. In this section, we focus on the fluctuation of the ETF's index

Table 4
The effect of ETF arbitrage and ETF's information share on the predictive relationship between decomposed ETF flows and index's total volatility

	$RV^{1min}_{i,t+1}$	$RV^{5min}_{i,t+1}$
$Flow^{backward}_{i,t}$	1.105	1.124
	(1.17)	(1.03)
$Flow^{forward}_{i,t}$	3.390***	3.893***
	(4.62)	(4.88)
$Arb_{i,t}$	2.978	3.059
	(1.45)	(1.29)
$Arb_{i,t} \times Flow^{backward}_{i,t}$	0.781	1.234
	(0.56)	(0.68)
$Arb_{i,t} \times Flow^{forward}_{i,t}$	4.246***	5.574***
	(3.97)	(4.13)
$IS_{i,t}$	0.427	0.486
	(1.12)	(0.99)
$IS_{i,t} \times Flow^{backward}_{i,t}$	0.215	0.178
	(1.02)	(0.86)
$IS_{i,t} \times Flow^{forward}_{i,t}$	0.176	0.204
	(1.43)	(1.55)
$Illq_{i,t+1}$	4.120***	4.683***
	(3.04)	(3.07)
$Volume_{i,t+1}$	2.825***	2.546***
	(11.07)	(9.05)
$Mk_{i,t+1}$	−7.436***	−7.822***
	(−3.62)	(−3.43)
Constant	−0.102	−0.197
	(−0.64)	(0.69)
First lag of dependent variable	0.182***	0.178***
	(30.96)	(30.08)
ETF fixed effect	Yes	Yes
Month fixed effect	Yes	Yes
Adj. R^2	0.27	0.26

The table extends Panel B in Table 3 by including six more variables: ETF arbitrage $Arb_{i,t}$, ETF's information share $IS_{i,t}$ and the four interaction terms, $Arb_{i,t} \times Flow^{backward}_{i,t}$, $Arb_{i,t} \times Flow^{forward}_{i,t}$, $IS_{i,t} \times Flow^{backward}_{i,t}$ and $IS_{i,t} \times Flow^{forward}_{i,t}$. The t-statistics are reported in parentheses. *** indicate significance at the 1, 5 and 10 percent levels, respectively.

driven by the variation of fundamental values. To quantify the fundamental volatility of the index, we apply two approaches. First, we conduct a BN decomposition with an autoregressive (AR) model, and second, we employ Hasbrouck's (1995) framework with the VECM.[16]

[16] BN decomposition is a widely used approach for efficient price determination and is also adopted by Hasbrouck (1993), Boehmer and Kelley (2009), Lee *et al*. (2016), and Wang and Yang (2017). Meanwhile, the approach of Hasbrouck (1995) is also quite famous in the microstructure literature that wants to resolve the contributions to a security's random-walk variance.

BN decomposition can be conducted through a vector autoregressive (VAR) system when one needs to resolve the contributions to a security's random walk variance. However, a univariate AR model is sufficient if one is interested in the random-walk variance itself, foregoing the component attributions, which is the case here. For this reason, we apply an AR model for BN decomposition.[17] Specifically, for each trading day, we regress the log return of an index in a 1-min sampling frequency over 30 lags: $r_\tau = \sum_{k=1}^{30} A_k r_{\tau-k} + \varepsilon_\tau$, which generates estimates $\Phi(1) = 1 - \sum_{k=1}^{30} \hat{A}_k$ and residuals $\hat{\varepsilon_\tau}$. According to BN, $\Phi(1)^{-1}\hat{\varepsilon_\tau}$ captures the permanent component of price shocks to the index while its variance $\Phi^{-2}\text{Var}(\hat{\varepsilon_\tau})$ proxies the fundamental volatility of the index, named as FV^{BN}.

Hasbrouck (1995) demonstrates that the (log) price of an index p_τ consists of the efficient price m_τ, which reflects the fundamental value, and transitory pricing error s_τ, which impounds various microstructure effects, i.e., $p_\tau = m_\tau + s_\tau$. The efficient price m_τ is not observable in the market but is shared with its ETFs. It follows a random walk that is driven by the new information about the fundamentals: $m_\tau = m_{\tau-1} + u_\tau$, where u_τ represents the arrival of new information at time τ and captures efficient price innovations. The variance in efficient price innovations, which is the variance of u_τ, proxies for the fundamental volatility, named as FV^{HAS}. Following Hasbrouck (1995, 2002), we apply a bivariate VECM for estimating FV^{HAS}. The VECM is processed using the data of the index and ETF at a 1-min frequency with 30 lags.

After constructing the series of FV^{BN} and FV^{HAS} for each ETF, we apply them to the panel regressions (2)–(4).[18] The dependent variable $Y_{i,t+1}$ now becomes $FV^{BN}_{i,t+1}$ and $FV^{HAS}_{i,t+1}$ instead of the two measures of the index's total volatility, the other settings remaining unchanged. The results are reported in Table 5.

In Panel A, where the undecomposed ETF flows are applied, it is shown that the coefficients of ETF flows are highly significant and positive in both columns. This suggests that ETF flows are correlated with the next day's variation of the underlying index driven by shocks to the fundamental value. ETF flows' capability of predicting fundamental volatility implies that ETF flows possibly contain APs' reaction to new information. Considering the two distinct roles of APs implied by ETF flows, market making versus trading on their own accounts, the latter compared to the former is apparently more likely

[17] We also conduct the BN decomposition through a VAR system and no qualitative difference is found.

[18] Again, we also conducted the time-series estimation for each of the 70 ETFs. No qualitative change against our major findings is detected. Results of time-series exercises are available upon request.

Table 5
The predictive relationship between (decomposed) ETF flows and index's fundamental volatility and the effects of ETF arbitrage and ETF's information share on these relationships

Panel A: The effect of undecomposed ETF flows

	$FV^{BN}_{i,t+1}$	$FV^{HAS}_{i,t+1}$
$Flow_{i,t}$	1.097***	1.134***
	(5.83)	(5.57)
Controls, constant and lagged dep.	Yes	Yes
ETF fixed effect and Month fixed effect	Yes	Yes
Adj. R^2	0.20	0.18

Panel B: The effects of decomposed ETF flows

	$FV^{BN}_{i,t+1}$	$FV^{HAS}_{i,t+1}$
$Flow^{backward}_{i,t}$	0.207	0.176
	(1.28)	(1.01)
$Flow^{forward}_{i,t}$	1.360***	1.494***
	(6.53)	(6.18)
Controls, constant and lagged dep.	Yes	Yes
ETF fixed effect and Month fixed effect	Yes	Yes
Adj. R^2	0.20	0.18

Panel C: The influences of ETF arbitrage and ETF's information share

	$FV^{BN}_{i,t+1}$	$FV^{HAS}_{i,t+1}$
$Flow^{backward}_{i,t}$	0.162	0.148
	(1.05)	(0.86)
$Flow^{forward}_{i,t}$	1.130***	1.243***
	(5.56)	(5.37)
$Arb_{i,t}$	15.515	15.974
	(1.27)	(1.06)
$Arb_{i,t} \times Flow^{backward}_{i,t}$	0.165	0.129
	(0.23)	(0.15)
$Arb_{i,t} \times Flow^{forward}_{i,t}$	0.338	0.147
	(0.72)	(0.26)
$IS_{i,t}$	3.247***	3.407**
	(2.82)	(2.50)
$IS_{i,t} \times Flow^{backward}_{i,t}$	−0.034	0.013
	(−0.56)	(0.20)
$IS_{i,t} \times Flow^{forward}_{i,t}$	0.145***	0.155***
	(3.63)	(3.45)
Controls, constant and lagged dep.	Yes	Yes
ETF fixed effect and Month fixed effect	Yes	Yes
Adj. R^2	0.21	0.19

Panel A reports the results of the panel regression of index's fundamental volatility on day $t+1$ against the undecomposed ETF flows $Flow_{i,t}$ on day t. Dependent variables $FV^{BN}_{i,t+1}$ and $FV^{HAS}_{i,t+1}$ are the fundamental volatility measures estimated by BN decomposition and Hasbrouck's (1995) model, respectively, on day $t+1$. Other settings are the same as Table 3. In Panel B, $Flow_{i,t}$, is replaced by backward-looking flows $Flow^{backward}_{i,t}$ and forward-looking flows $Flow^{forward}_{i,t}$. Panel C extends Panel B by including six more variables: ETF arbitrage $Arb_{i,t}$, ETF's information share $IS_{i,t}$ and their four interaction terms, $Arb_{i,t} \times Flow^{backward}_{i,t}$ and $Arb_{i,t} \times Flow^{forward}_{i,t}$, $IS_{i,t} \times Flow^{backward}_{i,t}$ and $IS_{i,t} \times Flow^{forward}_{i,t}$. The sample period starts in January 2015 and ends in December 2017. The t-statistics are reported in parentheses. *** and ** indicate significance at the 1, 5 and 10 percent levels, respectively.

to be informative because it is possible that APs react to new information for information profit (Xu *et al*., 2018). To confirm this, we replace $Flow_{i,t}$ by $Flow_{i,t}^{backward}$ and $Flow_{i,t}^{forward}$ and then replicate the regression. The results are given in Panel B of Table 5.

As expected, Panel B shows that the coefficients of the backward-looking flow component are insignificant in both columns, while those of the forward-looking component are highly significant and positive at the 1 percent level. Taking $FV_{i,t+1}^{BN}$for example, we find that $FV_{i,t+1}^{BN}$ increases by 22 percent of its standard deviation for every standard deviation increment in $Flow_{i,t}^{forward}$, whereas such increase for $Flow_{i,t}^{backward}$ is only 2 percent.[19] This confirms that APs' role as market makers is not responsible for the fundamental fluctuation of the index. On the other hand, the excess ETF shares created/redeemed on APs' own accounts can significantly increase the next day's fluctuation of the fundamental value of the index.

In Section 3, we argued that arbitrage opportunity intensifies the predictive power of the forward-looking flow component in the index's total volatility because sizable ETF mispricing stimulates APs to create/redeem more ETF shares for arbitrage profit and thus adds additional power for volatility prediction. However, it is a different story here. Because arbitrageurs do not necessarily possess information, arbitrage should not be closely correlated with fundamental volatility. On the other hand, an ETF's information share increases with information-driven ETF flows (Xu *et al*., 2018), which makes a greater *IS* imply more information-based creation/redemption of ETF shares. Given the mechanism of information contagion from ETFs to the underlying index (Xu and Yin, 2017a,b), an ETF's information share is likely to affect the fundamental volatility of the index through motivating more information-driven ETF flows. To confirm these inferences, we replicate regression (4) with the dependent variables being $FV_{i,t+1}^{BN}$ and $FV_{i,t+1}^{HAS}$. The results are reported in Panel C of Table 5.

Apparently, as we can see from Panel C, ETF arbitrage itself is irrelevant to the fundamental volatility of the index, and its role in affecting the relationships between the two flow components and the fundamental volatility of the index is also inconclusive. Nevertheless, an ETF's information share not only increases the index's fundamental volatility by itself but also enhances the predictive power of the forward-looking flow component on the index's fundamental volatility. These observations are highly consistent with our inferences. Moreover, due to the passive role of backward-looking flows in their response to demand, neither ETF arbitrage nor ETF's information share can significantly influence its impact on volatility. Other major observations, including unreported controls, in Table 5 are largely consistent with those in Tables 3 and 4.

[19] $1.360 \times \frac{0.021}{0.131} = 0.218; 0.207 \times \frac{0.015}{0.131} = 0.024.$

6. Granger causality between ETF flows and index volatility by VAR model

Having confirmed the correlation between ETF flows and index volatility in the context of the predictive relationships between the two flow components and the two types of index volatility, one may be interested in a further question: Are these relationships one-way or two-way? To answer this question, we employ the following VAR model between ETF flows and the two types of index volatility, which leads to Granger causality analysis:

$$Y_{i,t} = \alpha_1 + \sum_{k=1}^{p} \rho_{1,k} Y_{i,t-k} + \sum_{m=1}^{p} \pi_{1,m} Flow_{i,t-m} + \gamma_1^j Controls_{i,t}^j + e_{i,t}, \tag{5}$$

$$Flow_{i,t} = \alpha_2 + \sum_{k=1}^{p} \rho_{2,k} Y_{i,t-k} + \sum_{m=1}^{p} \pi_{2,m} Flow_{i,t-m} + \gamma_2^j Controls_{i,t}^j + v_{i,t} \tag{6}$$

Moreover, we extend to the multivariate VAR model with the two flow components:

$$\begin{aligned} Y_{i,t} = \alpha_1 &+ \sum_{k=1}^{p} \rho_{1,k} Y_{i,t-k} + \sum_{m=1}^{p} \pi_{1,m} Flow_{i,t-m}^{backward} + \sum_{n=1}^{p} \varphi_{1,n} Flow_{i,t-n}^{forward} \\ &+ \gamma_1^j Controls_{i,t}^j + e_{i,t}, \end{aligned} \tag{7}$$

$$\begin{aligned} Flow_{i,t}^{backward} &= \alpha_2 + \sum_{k=1}^{p} \rho_{2,k} Y_{i,t-k} + \sum_{m=1}^{p} \pi_{2,m} Flow_{i,t-m}^{backward} \\ &\quad + \sum_{n=1}^{p} \varphi_{2,n} Flow_{i,t-n}^{forward} + \gamma_2^j Controls_{i,t}^j + v_{i,t}, \end{aligned} \tag{8}$$

$$\begin{aligned} Flow_{i,t}^{forward} &= \alpha_3 + \sum_{k=1}^{p} \rho_{3,k} Y_{i,t-k} + \sum_{m=1}^{p} \pi_{3,m} Flow_{i,t-m}^{backward} \\ &\quad + \sum_{n=1}^{p} \varphi_{3,n} Flow_{i,t-n}^{forward} + \gamma_3^j Controls_{i,t}^j + \omega_{i,t}. \end{aligned} \tag{9}$$

We include two lags in the model, namely, $p = 2$.[20] The VAR estimations in this section are conducted based on time-series analysis rather than panel analysis. Therefore, regressions (5–9) are estimated for each of the 70 ETFs one by one, and then the results are averaged across all 70 ETFs. The average results are reported in Table 6, where Panel A (B) refers to the analysis on index's total volatility (fundamental volatility) with the undecomposed ETF

[20] Alternatively, we also applied more lags, $p = 5$, to the model. Including more lags does not lead to a qualitative change in our findings.

Table 6
Granger causality between (decomposed) ETF flows and the two types of index volatility with VAR estimations

Panel A: VAR estimation with undecomposed ETF flows and index's total volatility

	$RV_{i,t}^{1min}$	$Flow_{i,t}$
$RV_{i,t-1}^{1min}$	0.328***	0.012***
	(28.01)	(3.27)
$RV_{i,t-2}^{1min}$	0.090***	0.002
	(6.24)	(0.57)
$Flow_{i,t-1}$	1.746***	0.115***
	(4.67)	(3.32)
$Flow_{i,t-2}$	1.189***	0.103***
	(3.26)	(2.90)
Controls and constant	Yes	Yes
Adj. R^2	0.54	0.13

Panel B: VAR estimation with undecomposed ETF flows and index's fundamental volatility

	$FV_{i,t}^{BN}$	$Flow_{i,t}$
$FV_{i,t-1}^{BN}$	0.169***	0.042***
	(10.50)	(3.01)
$FV_{i,t-2}^{BN}$	0.107***	−0.0008
	(6.36)	(−0.59)
$Flow_{i,t-1}$	0.558***	0.118***
	(5.38)	(3.41)
$Flow_{i,t-2}$	0.230**	0.102***
	(2.17)	(2.95)
Controls and constant	Yes	Yes
Adj. R^2	0.30	0.13

Panel C: VAR estimation with decomposed ETF flow components and index's total volatility

	$RV_{i,t}^{1min}$	$Flow_{i,t}^{backward}$	$Flow_{i,t}^{forward}$
$RV_{i,t-1}^{1min}$	0.316***	−0.0002	0.012***
	(27.24)	(−0.80)	(3.23)
$RV_{i,t-2}^{1min}$	0.088***	0.0002	0.002
	(5.82)	(1.10)	(0.46)
$Flow_{i,t-1}^{backward}$	0.692	0.084**	0.281
	(1.15)	(2.35)	(1.19)
$Flow_{i,t-2}^{backward}$	0.812	0.097***	0.195
	(1.38)	(2.72)	(0.66)
$Flow_{i,t-1}^{forward}$	1.726***	0.002	0.127***
	(4.59)	(1.53)	(3.62)
$Flow_{i,t-2}^{forward}$	1.317***	0.001	0.108***
	(3.51)	(1.36)	(3.06)
Controls and constant	Yes	Yes	Yes
Adj. R^2	0.55	0.09	0.14

(continued)

Table 6 (continued)

Panel D: VAR estimation with decomposed ETF flow components and index's fundamental volatility

	$FV_{i,t}^{BN}$	$Flow_{i,t}^{backward}$	$Flow_{i,t}^{forward}$
$FV_{i,t-1}^{BN}$	0.165***	0.000	0.040***
	(10.08)	(0.81)	(2.73)
$FV_{i,t-2}^{BN}$	0.102***	−0.000	−0.0014
	(6.12)	(−0.56)	(−0.99)
$Flow_{i,t-1}^{backward}$	0.096	0.087**	0.252
	(0.52)	(2.48)	(0.89)
$Flow_{i,t-2}^{backward}$	0.260	0.100***	0.205
	(1.34)	(2.87)	(0.74)
$Flow_{i,t-1}^{forward}$	0.495***	0.002	0.121***
	(4.51)	(1.48)	(3.55)
$Flow_{i,t-2}^{forward}$	0.152	0.001	0.109***
	(1.39)	(1.20)	(3.06)
Controls and constant	Yes	Yes	Yes
Adj. R^2	0.30	0.10	0.14

Panels A and B report the outcome of the VAR estimations between the undecomposed ETF flows $Flow_{i,t}$ and the realised variance of 1-min index returns $RV_{i,t}^{1min}$, and between $Flow_{i,t}$ and the index's fundamental volatility estimated by BN decomposition $FV_{i,t+1}^{BN}$, respectively. Panel C (D) corresponds to Panel A (B) but uses the two decomposed ETF flows $Flow_{i,t}^{backward}$ and $Flow_{i,t}^{forward}$ instead of $Flow_{i,t}$. The regressions are run for each of the 70 sample ETFs one by one. The sample period of an ETF starts in January 2015 and ends in December 2017. The table reports the estimated coefficients, averaged across all ETFs. Average t-statistics are in parentheses. *** and ** indicate the overall significance at the 1, 5 and 10 percent levels, respectively.

flows, and Panel C (D) refers to the analysis on the index's total volatility (fundamental volatility) with the two decomposed components of ETF flows. For reasons of space, we only report the results for the realised variance of 1-min index returns $RV_{i,t}^{1min}$ as the proxy for the index's total volatility and $FV_{i,t}^{BN}$ as the proxy for index's fundamental volatility.[21]

In Panels A and B, the coefficients of the first two lagged terms of ETF flows are on average significant at the 1 percent level in the regressions of $RV_{i,t}^{1min}$ and $FV_{i,t}^{BN}$. On the other hand, the coefficients of $RV_{i,t-1}^{1min}$ and $FV_{i,t-1}^{BN}$ are also significant at the 1 percent level in the regression of $Flow_{i,t}$. In the VAR system, a Granger causality test in nature examines the joint significance of the effects of the lagged terms of the explanatory variables, i.e., as long as one of those lagged terms is significant, Granger causality holds. Thus, it is safe to conclude from Table 6 that a two-way Granger causality exists between ETF flows and the two types of index volatility; that is, more ETF share creation/redemption

[21] Results for other volatility proxies are highly consistent with those presented in Table 6. They are available from the authors upon request.

Granger-causes the index to be more (fundamentally) volatile, and a more (fundamentally) volatile index also Granger-causes more ETF shares created/redeemed.

From Panels C and D, no evidence is found for a lead-lag relationship between backward-looking flows and either type of index volatility. On the other hand, we provide strong evidence supporting a two-way Granger causality between the forward-looking flows and the two types of volatility in the sense that both $Flow_{i,t-1}^{forward}$ and $Flow_{i,t-2}^{forward}$ are significant in explaining $RV_{i,t}^{1min}$ and $FV_{i,t}^{BN}$, whose first lagged terms are also significant in explaining $Flow_{i,t}^{forward}$. Recalling the two-way Granger causality between the undecomposed ETF flows and the two types of index volatility, one can conclude that such a two-way relationship is largely because of the role played by the forward-looking flows. This suggests that not only the APs' excess share creation/redemption beyond their role as market makers can predict the future (fundamental) volatility of the index; it is also the case that a (fundamentally) volatile index may change APs' potential profits for their trading and thus vary their decision-making for the excess share creation/redemption.

While we provide evidence that the channel connecting ETF flows, forward-looking flows in particular, and index volatility is a two-way contagion, we like to note a caveat here that Granger causality is different from actual causality, as it only implies the time-serial 'causal' effect by indicating the time sequence of occurrence. Although Granger causality does not equal actual casualty, this finding still provides additional evidence of strong interaction and close correlation between ETF flows and the two types of index volatility.

7. Robustness checks

This section summarises the robustness tests for the analyses throughout the previous sections. First, fundamental volatility measures including both FV^{BN} and FV^{HAS} are constructed based on the intraday data at a 1-min frequency with 30 lags in the AR and VECM frames. For robustness, we also apply 10, 20 and 60 lags in addition to the setting of 30 lags for estimating FV^{BN} and FV^{HAS}, and then replicate the analysis. Second, the time-trend is not a significant issue over the 2015–2017 period for all variables, as we have noted previously. For robustness, we also detrend all the regression variables by obtaining the residuals from regressing each variable against yearly dummies and a linear time trend, following Chordia *et al.* (2006), and then replicate all analyses. Third, in additional to the panel analysis in Sections 4 and 5, the time-series analysis on each of the 70 ETFs are also conducted one by one. Fourth, in addition to Hasbrouck's (1995) information share, we also employ the modified information share developed by Lien and Shrestha (2009) for robustness. Fifth, the VAR models in the Granger causality section are specified with two lags. Further experiments with five lags are also conducted.

Fortunately, none of the above checks leads to a qualitative change to those results reported in the previous tables, and it thereby confirms that our findings are very robust. The results of these exercises are not reported but they are available upon request.

8. Concluding remarks

This paper sheds novel insights on the decomposed components of ETF flows and their associations with the total volatility and fundamental volatility of the underlying index. By providing evidence from the emerging market in China, we reveal that ETF flows significantly increase both types of index volatility on the next trading day, and such predictive power of ETF flows is largely because of the forward-looking flow component, which indicates APs' additional share creation/redemption activities beyond its response to market demand. However, the effects of backward-looking flows, which gauge APs' passive reaction to market demand in predicting the two types of index volatility, are rather limited. Furthermore, with a greater end-of-day ETF arbitrage, the forward-looking flows exert a greater effect in predicting the total volatility, but not the fundamental volatility, of the ETF's index. On the other hand, as an ETF shares more new information, the predictive power of the forward-looking flows is enhanced in the fundamental volatility, but not the total volatility, of the ETF's index. Finally, we show that the relationships between ETF flows, the forward-looking flow component in particular, and the two types of index volatility are a two-way contagion.

Our findings also point to the fact that the passively managed ETFs are not really passive, which is probably because of the dominance of the active trading activities of APs for their own advantages over their passive role as market makers in affecting the fluctuation of the whole index. This paper thus places a caveat for not only ETF investors but also policy makers not to underestimate the power of APs and that they are not passive players. APs play the game actively and effectively.

While the current study explores the tight correlation between ETF flows and the next day's index volatility, we have not addressed the efficiency of the index and its ETFs. For example, we have not answered the questions such as whether ETF flows and their specific components are able to predict the future price efficiency of the underlying index, and, if yes, whether they are a reflection of private or public information. These research questions remain to be answered in future research.

References

Amihud, Y., 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Baillie, R. T., G. G. Booth, Y. Tse, and T. Zabotina, 2002, Price discovery and common factor models, *Journal of Financial Markets* 5, 309–321.

Ben-David, I., F. Franzoni, and R. Moussawi, 2017, Do ETFs increase volatility?, *The Journal of Finance* 47, 2015–2034.
Beveridge, S., and C. R. Nelson, 1981, A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the 'business cycle', *Journal of Monetary Economics* 7, 151–174.
Boehmer, E., and E. K. Kelley, 2009, Institutional investors and the informational efficiency of prices, *Review of Financial Studies* 22, 3563–3594.
Broman, M. S., and P. Shum, 2018, Relative liquidity, fund flows and short-term demand: evidence from exchange-traded funds, *The Financial Review* 53, 87–115.
Brown, D. C., S. Davies, and M. Ringgenberg, 2018, ETF arbitrage and return predictability. Available at: https://ssrn.com/abstract = 2872414.
Chen, X., Y. Liu, and T. Zeng, 2017, Does the $T + 1$ rule really reduce speculation? Evidence from Chinese Stock Index ETF, *Accounting and Finance* 57, 1287–1313.
Chordia, T., R. Roll, and A. Subrahmanyam, 2001, Market liquidity and trading activity, *The Journal of Finance* 56, 501–530.
Chordia, T., A. Sarkar, and A. Subrahmanyam, 2005, An empirical analysis of stock and bond market liquidity, *The Review of Financial Studies* 18, 85–129.
Chordia, T., S. Huh, and A. Subrahmanyam, 2006, The cross-section of expected trading activity, *The Review of Financial Studies* 20, 709–740.
Chou, R. Y., 1988, Volatility persistence and stock valuations: some empirical evidence using GARCH, *Journal of Applied Econometrics* 3, 274–294.
Clark, P. K., 1973, A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* 41, 135–155.
Clifford, C. P., J. A. Fulkerson, and B. D. Jordan, 2014, What drives ETF flows?, *Financial Review* 49, 619–642.
Granger, C. W. J., and O. Morgenstern, 1963, Spectral analysis of New York stock market prices, *Kyklos* 16, 1–27.
Greenwood, R., 2005, Short- and long-term demand curves for stocks: theory and evidence on the dynamics of arbitrage, *Journal of Financial Economics* 75, 607–649.
Hasbrouck, J., 1993, Assessing the quality of a security market: a new approach to transaction-cost measurement, *The Review of Financial Studies* 6, 191–212.
Hasbrouck, J., 1995, One security, many markets: determining the contributions to price discovery, *The Journal of Finance* 50, 1175–1199.
Hasbrouck, J., 2002, Stalking the 'efficient price' in market microstructure specifications: an overview, *Journal of Financial Markets* 5, 329–339.
Hua, R., Q. Liu, and Y. Tse, 2016, Extended trading in Chinese index markets: informed or uninformed?, *Pacific-Basin Finance Journal* 36, 112–122.
Karpoff, J. M., 1987, The relation between price changes and trading volume: a survey, *Journal of Financial and Quantitative Analysis* 22, 109–126.
Krause, T. A., S. Ehsani, and D. D. Lien, 2014, Exchange traded funds, liquidity, and market volatility, *Applied Financial Economics* 24, 1617–1630.
Kyle, A. S., and W. Xiong, 2001, Contagion as a wealth effect, *The Journal of Finance* 56, 1401–1440.
Lamoureux, C., and W. D. Lastrapes, 1990, Persistence in variance, structural change, and the GARCH model, *Journal of Business and Economic Statistics* 8, 225–234.
Lee, C., K. H. Chung, and S. Yang, 2016, Corporate governance and the informational efficiency of prices, *Financial Management* 45, 239–260.
Lettau, M., and A. Madhavan, 2016, ETF 101 for economists, Working paper (University of California at Berkeley).
Lien, D., and K. Shrestha, 2009, A new information share measure, *Journal of Futures Markets* 29, 377–395.

Madhavan, A., 2016, *Exchange-Traded Funds and the New Dynamics of Investing* (Oxford University Press, Oxford).
Madhavan, A., U. Marchioni, W. Li, and D. Y. Du, 2014, Equity ETFs versus index futures: a comparison for fully funded investors, *Journal of Index Investing* 5, 66–75.
McKenzie, M. D., and R. W. Faff, 2003, The determinants of conditional autocorrelation in stock returns, *Journal of Financial Research* 26, 259–274.
Petajisto, A., 2016, Inefficiencies in the pricing of exchange-traded funds, Working paper. Available at: https://ssrn.com/abstract=2000336.
Staer, A., 2017, Fund flows and underlying returns: the case of ETFs, *International Journal of Business* 22, 275–304.
Stratmann, T., and J. W. Welborn, 2012, Exchange-traded funds, fails-to-deliver, and market volatility, Working paper. Available at: http://ssrn.com/abstract=2183251.
Wang, K. Q., and J. Xu, 2012, Market volatility and momentum, *Journal of Empirical Finance* 30, 79–91.
Wang, J., and M. Yang, 2017, Conditional volatility persistence, Working Paper (University of Technology Sydney).
Weber, P., and B. Rosenow, 2006, Large stock price changes: volume or liquidity?, *Quantitative Finance* 6, 7–14.
Xu, L., and X. Yin, 2017a, Does ETF trading affect the efficiency of the underlying index?, *International Review of Financial Analysis* 51, 82–101.
Xu, L., and X. Yin, 2017b, Exchange traded funds and stock market volatility, *International Review of Finance* 17, 525–560.
Xu, L., X. Yin, and J. Zhao, 2018, Are authorized participants of exchange-traded funds informed traders?, Working Paper. Available at: http://ssrn.com/abstract=3221852.