Journal of International Money and Finance 149 (2024) 103194

Contents lists available at ScienceDirect

Journal of International Money and Finance

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

ETFs and tail dependence: Evidence from Chinese stock market

Wei Ning [a], Jiahua Zhao [b,*], Fuwei Jiang [c]

[a] School of Finance, Southwestern University of Finance and Economics, China
[b] School of Finance, Central University of Finance and Economics, China
[c] Department of Finance at School of Economics, Wang Yanan Institute for Studies in Economics, Xiamen University, China

ARTICLE INFO

Keywords:
Exchange-Traded Funds (ETFs)
Tail Dependence
Arbitrage Activity
Financial Network

ABSTRACT

Using Chinese A-share market data, we empirically examine the impact of exchange-traded funds (ETFs) on tail dependence of the underlying securities. Our results show that ETFs can increase the tail dependence of stocks in their basket, showing that the average tail dependence of a stock is higher when it has stronger ETF holding similarity with other stocks. We investigate the role of arbitrage activity and find that ETF holding similarity increases stocks' ETF arbitrage activity. This effect is primarily on discount arbitrage rather than premium arbitrage, which leads to higher tail dependence among stocks. Alongside propagating demand shocks from the ETF market to underlying securities, ETFs also propagate tail event shock from one stock to other stocks in their baskets. Additionally, arbitrage activities through ETFs add a new layer of non-fundamental tail dependence to the underlying securities. Unlike mutual funds, ETFs lead to more frequent and idiosyncratic tail risk contagion among underlying securities. Our study sheds light on how ETFs provide new channels for risk contagion among underlying securities in emerging markets.

1. Introduction

Investing in financial markets has increasingly shifted towards indexing. As the most significant financial innovation in recent years, exchange-traded funds (ETFs) provide investors with a convenient, low-cost, and highly liquid tool for index investing (Ben-David et al., 2018). Index investing generates great demand for ETFs, making it witness tremendous growth. Since introduced in the early 1990 s, the total assets under management (AUM) of ETFs have climbed to approximately $12 trillion by the end of 2023, with a compound annual growth rate (CAGR) of 22.16 % over the past two decades. In China, index investing has become the trend in the wealth management industry. By the end of 2023, China had 889 ETF products, with assets under management over 2 trillion RMB and total net flow over 0.5 trillion RMB.(See Fig. 1).

Studies reveal that shifting to index investing may affect financial stability through its impacts on asset management industry concentration, and co-movement of asset return and liquidity (Anadu et al., 2020). Since ETFs will by all measures play a crucial role in the future of investing, it is important to understand their economic impact on financial market and underlying securities. The literature on the economic consequences of ETFs has grown rapidly alongside the development of the ETF market. Majority of studies employ proxies for ETF ownership to analyze the economic impact of ETFs on stock volatility (Ben-David et al., 2018), return predictability (Brown et al., 2021), informational efficiency (Glosten et al., 2021), and real investment (Antoniou et al., 2023). Like other

* Corresponding author at: School of Finance, Central University of Finance and Economics, Changping District, Shahe University Park, Beijing 102206, China.

E-mail addresses: ningwei@swufe.edu.cn (W. Ning), jiahua0714@163.com (J. Zhao), fwjiang@xmu.edu.cn (F. Jiang).

https://doi.org/10.1016/j.jimonfin.2024.103194

Available online 12 September 2024

passive funds, ETFs hold underlying securities that are included in their benchmark index. With the rapid growth of ETFs, financial assets are progressively integrated into common portfolios, leading to heightened interconnectedness among these assets. One may wonder whether the rising of interconnectedness has unintended consequences for the securities in ETFs' baskets.

This study explores the impact of ETFs on the tail dependence of their underlying securities. Existing empirical studies have examined the influence of financial products on asset prices (Coval and Stafford, 2007; Chernenko and Sunderam, 2020; Feinstein and Halaj, 2023), with a primary focus on the risk contagion from these products to the underlying securities. For instance, Ben-David et al. (2018) find that the liquidity shocks of ETFs can propagate to the underlying securities through the arbitrage channel, and ETFs may increase the nonfundamental volatility of the securities in their baskets. Besides this channel, it remains unclear whether an increase in interconnectedness through ETFs provides a new channel for risk contagion between underlying securities, and further in-depth study is needed. Moreover, interconnectedness is an inherent feature of the modern financial system (Feinstein and Halaj, 2023), and there has been an upsurge in empirical and theoretical work on the relationship between interconnectedness and financial stability since the financial crisis of 2007–08 (Raddant and Kenett, 2021; Egger et al., 2023; Hsiao and Chiu, 2024). While previous studies provide strong evidence explaining how interconnectedness among financial institutions or financial products provide new channels for risk contagion (Choi et al., 2020; Duarte and Eisenbach, 2021; Glasserman and Young, 2016), less attention has been given to the interconnectedness among financial assets resulting from common asset holding.

The Chinese A-share market presents several shared and distinctive characteristics compared to mature markets such as the U.S., making it an ideal laboratory for testing how ETFs increase tail dependence among underlying securities. First, China stands as the largest emerging market in the world, and the Chinese A-share market attracts significant international attention. In recent years, foreign institutional investors have increased their investments in the Chinese market, while ETFs become a crucial tool for their index investing. For instance, in the first quarter of 2024, Qualified Foreign Institutional Investors held 744 A-share market stocks, amounting to a total market value exceeding 150 billion RMB. Over half of these investors also held ETF products. Notably, among the top ten ETF shareholders were seventeen foreign institutions, including Barclays Bank, UBS Group, Merrill Lynch International, and Morgan Stanley International.

Second, the rise of ETFs in China reflects several features of the global ETF market. Globally, 74.24 % of ETFs are equity-based, a figure that increases to 84.39 % in China. Broad-based ETFs comprise over 50 % of equity ETFs in both markets, requiring substantial holdings of benchmark index components by ETFs in each market. Additionally, both global and Chinese ETF markets emphasize innovation. Numerous ETFs focusing on ESG, energy transition, and emerging industries have been launched in both markets, resulting in ETFs holding stocks with similar characteristics.

Third, The Chinese A-share market presents several distinctive characteristics compared to mature markets, helping us better explore the impact of ETFs on tail dependence among underlying securities. Among market environment, China's stock market has a larger proportion of individual investors. Those small retail investors have low financial literacy, exhibit behavioral biases, and the irrational investment behavior of such investors is more likely to lead to market instability (Jones et al., 2023; Kelley and Tetlock, 2017; Tan et al., 2023). In terms of trading mechanisms, both Chinese stock and ETFs market are subject to the T+1 rule, which requires investors to hold the asset for at least 1 day before selling (Guo et al., 2012), making it easier for us to separate the secondary market arbitrage activities (Wu and Zhu, 2023), and identifying arbitrage channel through primary market and the secondary market

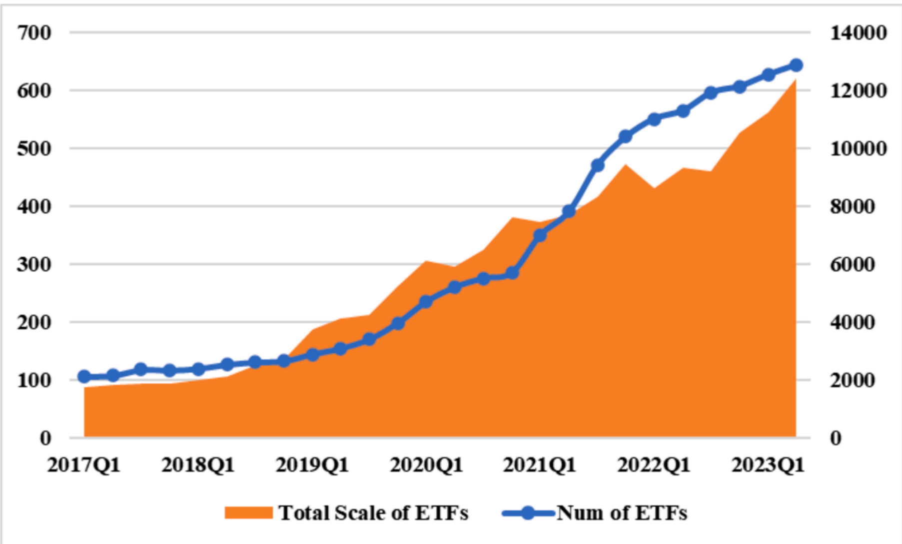


Fig. 1. The Scale and Number of Chinese ETFs. Notes: This figure illustrates the time series trend of the total scale and number of ETFs in the Chinese market during the sample period. The yellow area chart represents the total scale of ETFs in the Chinese market, measured in hundred millions of yuan, while the blue line chart indicates the number of ETF funds in the market. The horizontal axis shows the quarterly time labels, the left vertical axis corresponds to the number of funds, and the right vertical axis represents the scale.

of ETFs. At the product level, China issues many thematic ETFs (e.g. dividend ETFs, value ETFs, SoE ETFs, etc.), making ETF asset holdings exhibit a higher degree of homogeneity. Index providers often issue multiple ETFs within the same industry or theme, leading to highly similar ETFs holding between stocks, while these ETFs hold securities with shared characteristics but no fundamental relationship.

To test whether ETFs increase tail dependence among underlying securities, we calculate the pairwise ETF holding similarity based on the ETF holding weight matrix of underlying securities. We then construct a similarity matrix with security code as rows(columns) and similarities as elements. Meanwhile, drawing inspiration from Wang et al. (2021), we employ a dynamic mixture Clayton-survival Gumbel copula model to construct the tail dependence matrix of underlying securities. We take the constituent stocks of the CSI 300 Index as a subsample to provide the basic visualization intuition. By converting those two matrices into heatmaps, we observe that both similarities and tail dependence move to a wide range of red colors from 2017 to 2022, signaling increasing interconnectedness through ETFs holding and tail dependence among the underlying securities of CSI 300 Index.

Motivated by the visualization intuition, we calculate the average of the similarity proxy and tail dependence measurement at the stock level. Subsequently, we conduct a series of analyses using a panel of stock-quarterly observations between 2017 and 2023. Our findings indicate that ETFs will increase tail dependence among underlying securities, showing that the average tail dependence of an individual stock is higher when it exhibits stronger ETF holding similarity with other stocks. Our results are robust at least in the 99 % confidence interval after conducting control variables at different level, and including a range of fixed effect (e.g., firm industry, province). For economic significance, a one standard deviation increase in the interconnectedness can lead to tail dependence increasing by 39.95 % of their standard deviation.

To better explore the relationship between ETFs and tail dependence, as well as solve the endogeneity problem, we employ the difference-in-difference method to present daily-level evidence. We utilize regulatory penalty events for Chinese A-share stocks as quasi-natural experiments, where the 10 stocks with the highest ETF holding similarity to the penalized stocks prior to the penalty event are treatment group stocks, and the 10 stocks with the lowest ETF holding similarity to the penalized stocks are control group stocks. Our experiment design shows several desirable properties. The (cumulative) abnormal return of the three subsamples (penalized, treatment and control group) exhibits no significant pre-trends, and the regulatory penalty events are staggered over time. We estimate a difference-in-difference regression model in which we compare the cumulative abnormal return in treatment group stocks versus control group stocks before and after the regulatory penalty events. Our results indicate that, in comparison to control group stocks, the cumulative abnormal return for treatment group stocks will significantly decrease by 0.5 % after the regulatory penalty events.

Compared with traditional index funds, ETFs can simultaneously create and redeem shares in the primary market and trade shares in the secondary market. This unique trading mechanism facilitates the activities of authorized participants (APs), since they have the right but not the obligation to create or redeem ETF shares (Sommers, 2024). Recent studies argue that cross-market arbitrage activity could make markets more fragile and result in price crashes (Menkveld and Yueshen, 2019). Building on insights from literature exploring the role of arbitrage activity (Huang et al., 2015; Ben-David et al., 2018; Jiang et al., 2019; Fulkerson et al., 2022; Huang et al., 2021), we investigate the relationship between ETFs, arbitrage activity, and tail dependence. Our findings reveal that ETF holding similarity significantly increases stocks' ETFs arbitrage activity, and such enhancement is observed primarily in discount arbitrage instead of premium arbitrage, which leads to higher tail dependence among stocks. Daily-level difference-in-difference regression analyses indicate that, in comparison to control group stocks, the ETFs arbitrage activity for treatment group stocks significantly increases after regulatory penalty events.

While previous studies have discussed the economic impact of ETFs on underlying securities as a source of risk, our paper emphasizes the role of ETFs in propagating shocks from one constituent stock to others. Therefore, we need to empirically test how our findings differ from existing studies. We first examine whether ETFs propagate stock-level shocks rather than originating shocks themselves. Following Ben-David et al. (2018), we use stock-level ETF ownership and ETF volatility to measure the risk pressure originating from ETFs.We then perform a regression analysis of tail dependence on ETF ownership and ETF volatility, controlling for other potential influences on tail dependence. The regression residuals are used as a measure of excess tail dependence, representing the portion of stock tail dependence that cannot be explained by ETF risk alone. If ETFs significantly impact excess tail dependence, it suggests that ETFs provide a new source of risk beyond their inherent risk. Our analysis reveals a substantial increase in stock excess tail dependence due to ETF interconnectedness. This finding indicates that ETFs not only contribute to individual stock risk but also facilitate risk contagion among stocks through their holdings.

Next, we explore whether ETFs propagate non-fundamental risk. Following literatures on stock price synchronicity (Chan and Hameed, 2006; An and Zhang, 2013; Chan and Chan, 2014), we construct two stock price synchronicity proxies and regress tail dependence on these proxies along with other variables that may affect stocks' tail dependence. Since stock price synchronicity reflects the extent to which market and industry fundamental information is incorporated into stock returns, we use the regression residuals as a proxy for non-fundamental tail dependence between stocks. Our results reveal that interconnectedness through ETFs can explain the portion of tail dependence among stocks that cannot be accounted for by market or industry fundamental information, implying that ETFs do propagate non-fundamental risk.

Finally, we examine the differences between ETFs and mutual funds in enhancing stocks' tail dependence. Mutual fund also holds many underlying financial assets, leading to tight connections between underlying securities. Unlike mutual funds, however, ETFs can be traded in real-time and redeemed continuously on the secondary market. Therefore, we expect that ETFs exhibit more frequent and individualized characteristics in amplifying stock tail dependencies compared to mutual fund. We employ quarterly mutual fund asset holding data and construct a mutual fund connection indicator. To assess whether ETFs demonstrate high-frequency variability in risk contagion, we create a market stress dummy variable along with two connection stock stress dummy variables. We regress tail dependence on ETFs and mutual fund connection proxies, as well as their interactions with the stress dummy. Our findings indicate that both ETFs and mutual funds increase tail dependence between stocks during market stress periods. However, during stock stress periods, only ETF interconnectedness enhances tail dependence between underlying securities. Overall, we find that compared to mutual funds, ETFs exhibit high-frequency risk contagion characteristics, amplifying the spread of idiosyncratic risks among stocks.

Our paper contributes to the literature along several aspects. First, we contribute to the growing debate on the consequences of ETFs. One strand of these studies constructs proxies for ETF ownership and discuss the economic effects of ETFs on stock volatility (Ben-David et al., 2018), return predictability (Brown et al., 2021), informational efficiency (Xu and Yin, 2017; Glosten et al., 2021), and real investment (Antoniou et al., 2023). Other studies attempt to understand the unintended consequences of ETFs on financial market (Box et al., 2019; Bhojraj et al., 2020; Ben-David et al., 2023). Although studies provide evidence on the effect of ETFs and other derivatives on the quality of the underlying securities' prices, they mostly focus on the direct impact of ETFs. In this study, we construct a ETF holding similarity proxy for underlying securities and focus on how ETFs as a risk contagion channel increase tail dependence, which broadens this strand of literature.

Our paper is closely related to empirical studies on ETFs and underlying securities' risk. Existing studies shows that shocks of ETFs can propagate to the underlying securities through different channels. For example, Ben-David et al. (2018) find that ETFs arbitrage activity increases nonfundamental volatility in underlying stocks. Da and Shive (2018) show that ETF arbitrage contributes to return co-movement. Pan and Zeng (2017) show that the intermarket liquidity mismatch in ETFs generates market instability. Israeli et al. (2017) find that ETFs ownership increases the bid-ask spread on stocks. Our results show that shocks of an individual stock can also propagate to other stocks in the same ETF basket, which provides more empirical evidence to comprehensively understand the impact of ETFs on underlying securities' risk.

Our results also add to the literature on tail dependence. We build upon prior studies that have explored tail dependence structures across various assets and markets (Ye et al., 2017; Naeem and Karim, 2021). By focusing on the causal relationship between ETFs and underlying securities' tail dependence, our study improves the understanding of tail dependence formation. Moreover, while previous studies mainly focus on how common asset holding of financial institutions or products brings higher tail dependence and provide new channel for risk contagion (Hautsch et al., 2015; Cont and Schaanning, 2019; Barucca et al., 2021), less attention has been paid to the tail dependence among financial assets. We contribute to tail dependence literature by providing empirical evidence that common asset holding by ETFs can also amplify the tail dependence of financial assets in their basket.

The remainder of the paper is structured as follows. Section 2 presents our theoretical framework. Section 3 shows our data, variable construction, and descriptive statistics. Section 4 examines the empirical relationship between ETFs networks and tail dependence. Section 5 explores the underlying channels of arbitrage activity, and Section 6 gives additional analysis and robustness check. Section 7 concludes.

2. Theoretical framework

We conjecture that the arbitrage between ETFs and underlying securities can propagate shocks from one constituent to other stocks in their baskets, resulting in higher tail dependence among underlying securities. We use the frameworks of Ben-David et al. (2018) and Capponi et al. (2020) to explain the economic channel that we wish to identify.

ETFs provide investors with highly liquid and low-cost tools for index investing. Their primary market creation/redemption and secondary market trading offer authorized participants opportunities for arbitrage activities. We first utilize the framework in Ben-David et al. (2018) to explain how high-frequency arbitrage trading can lead to tail dependence among underlying securities. When a tail event occurs for an ETF component, a fundamental shock hits the ETF holding the stock, and the ETF price adjusts immediately as the fundamental information reaches the secondary market first. However, the indicative optimized portfolio value (IOPV) remains temporarily fixed due to stale pricing. For example, stock exchanges in China refresh their IOPV estimates for ETF products every 15 seconds. Consequently, the fire-sale pressure in the secondary market leads to a discount between the ETF trading price and the IOPV. Arbitrageurs who observe a profit opportunity will buy the ETF at the undervalued price in the secondary market and redeem it for a basket of stocks at the higher IOPV in the primary market, then sell the basket for an arbitrage gain. This selling pressure ultimately transmits the tail risk from the individual component to other constituent stocks in the ETF basket, thereby increasing tail dependence among the stocks.

The underlying premise of the above arbitrage activity is predicated on the stale pricing of IOPV and the real-time updating of ETF trading prices. However, due to trading latency and other constraints, arbitrage traders often encounter difficulties in completing arbitrage activities within the short timeframe of IOPV adjustment, raising concerns about the effectiveness of such high-frequency arbitrage activities. We posit that beyond the immediate adjustment period, arbitrage opportunities may also exist after the IOPV adjustment, eventually leading to increased tail dependence among component stocks.

We refer to the theoretical model in Capponi et al. (2020) to explain this mechanism based on the heterogeneity of investor microstructure. Capponi et al. (2020) develop a model of the feedback between mutual fund outflows and asset illiquidity. In this model, first movers anticipate and react to the final change in the fund's NAV, while second movers react only to the observed change in NAV. Due to advantages in information and trading capabilities, first movers act quickly and redeem fund shares before the fund begins to sell asset shares, thereby amplifying the asset fire-sales process, leading to larger price declines and spiraling redemptions. Given the significant heterogeneity among secondary market participants in ETFs, we utilize this framework to further explain the price dynamics when ETFs experience fundamental shocks. The presence of informed traders who can quickly react to fundamental information contributes to a feedback loop where their actions cause cascading effects on the prices of underlying stocks. This dynamic, where informed first movers initiate trades that influence subsequent market reactions, helps elucidate how ETF trading can lead to increased tail dependence among constituent stocks.

Specifically, we imagine a scenario involving two types of investors in the ETF market: first-mover investors with information and trading advantages, and second-mover investors who trade based solely on observed prices. When a tail event occurs for an ETF component stock, first movers immediately receive this information and sell ETFs in the secondary market to avoid potential risks, exerting selling pressure on the underlying stocks. Since the first movers sell ETF shares based on the expectation of adjusted IOPV, one can expect that the ETF price will realign with the IOPV after the issuer's adjustment. However, as second movers observe the declined ETF price, they will further sell their ETF holdings, leading to additional downward pressure on the ETF market price. This can cause the price to fall below the adjusted IOPV, creating a discount arbitrage opportunity. Arbitrageurs, as described in Ben-David et al. (2018), buy the undervalued ETF in the secondary market and redeem it in the primary market for a basket of stocks at the higher IOPV, subsequently selling these stocks for risk-free profits. Eventually, this process leads to a decline in the price of the underlying securities in the ETF, increasing tail dependence among the constituents. This chain reaction, initiated by informed first movers and perpetuated by second movers and arbitrageurs, illustrates how ETFs can propagate tail risks and amplify interconnectedness among component stocks.

In summary, our theoretical framework suggests that discount arbitrage arising from stale pricing and heterogeneity investor structures can propagate price pressure from individual constituent to other component stocks in ETFs, thereby increasing tail dependence among stocks.

3. Data, variables, and summary statistics

3.1. Data

3.1.1. ETFs data

We obtain ETFs return and holding data from the China Stock Market & Accounting Research (CSMAR) database. We restrict ETFs to equity- typed ETFs that have at least 50 % investment in China A-share market common stocks. Therefore, we exclude ETFs that are classified as bond, foreign equity, inverse and leveraged. Our sample period is from January 2017 to September 2023, during which the number of equity- typed ETFs in the Chinese market has risen rapidly from just over 100 to 687.

We retrieve daily ETF holdings data from CSMAR's ETFs Fund Subscription and Redemption List dataset. This dataset contains information on the conversion between ETFs shares and underlying securities in each trading day, including stock names and required number of shares. Given the high-frequency and time-varying nature of ETFs arbitrage activity, we choose this dataset rather than quarterly disclosed ETF holdings, which only reflects the ETFs stock holdings on the last trading day of the quarter. We merge this dataset with the previous day's closing price data of stocks to calculate the holding number of ETFs underlying securities, and further calculate the ETF holding weight of each stock. This procedure finally yields 1581 daily ETF holding similarity matrices, where the rows represent the stock codes, the columns represent the ETFs codes, and the matrix elements represent the ETF holding weight of stocks.

3.1.2. Stock data

Stock return and accounting data are from the intersection of the CSMAR and Wind from 2017 to 2023. Our sample includes firms listed on the Shanghai stock exchange and Shenzhen stock exchange that have share codes begin with 60, 30 or 00. We exclude stocks with average total market value less than 500 million to mitigate small size effect and exclude sample data with less than one year of issuance times to overcome new issue effect. To alleviate the effects of outliers, we winsorize all independent variables at the 0.5 % and 99.5 % levels. The final sample contains 64,688 firm-quarter observations.

We use data in the Chinese Listed Company Penalty Research Database in CSMAR to construct the quasi-natural experiment for causality identification. This dataset collects penalty information disclosed by stock exchanges and the China Securities Regulatory Commission (CSRC) from 1994, including violation dates, violation types, and penalties, and it is widely used in the study of financial regulation and risk contagion (He and Luo, 2018; Liu et al., 2022). We select penalty events with RelationshipID between P2402 and P2405 to ensure the direct relationship of regulatory penalties and stocks. and events with PunishmentTypeID between P2604 and P2607, to exclude minor forms of punishment such as criticisms, warnings, and denouncement. To prevent the interference of significant disclosure events, we exclude the penalty events with major announcements (e.g., financial statement disclosure, change of executives) released five trading days before the penalty announcement date. We finally get 5,351 regulatory penalty events, which involves 1,856 stocks.

3.2. Variable construction

3.2.1. Measuring interconnectedness through ETFs

Unlike mutual funds, which process share creation and redemptions in cash, ETF utilize a basket of stocks for their share transactions. Investors are required to provide a specified portfolio of stocks to the ETF issuer when creating ETF shares, and vice versa. This mechanism naturally enhances trading linkages among underlying securities. Furthermore, if two stocks are simultaneously held by multiple ETFs, we can anticipate a tighter connection being established between these stocks. Following this economic intuition, we use the similarity of ETFs' holding weight vectors of two stocks to compute the pairwise interconnectedness through ETFs. The interconnectedness of an individual stock is proxied as its average ETF holding similarity with other stocks. Specifically, the interconnectedness for of stock i in day t, $interconnect_{i,t}$, is calculated as:

$$interconnect_{i,t} = \frac{1}{N}\sum_{j=1}^{N} intersection_{ij,t} * cosinesimilarity_{ij,t} \tag{1}$$

Where $j \in \{1, \cdots, N\}$ is the set of stocks that established ETFs interconnectedness with stock i; We use a common measure $cosinesimilarity_{ij,t}$, which is the cosine similarity of stock i's and stock j's ETF holding weight vectors, to measure the ETF holding similarity. As a wildly used similarity measurement (Hanley and Hoberg, 2012; Girardi et al., 2021), this method measures the similarity by the cosine of the angle between two vectors, it automatically ignores elements that are both equal to 0 in two vectors, making the pairwise cosine similarity non-comparable. To address this problem, we multiply each cosine similarity by a multiplier $intersection_{ij,t}$, which is the length of the intersection of stock i's and stock j's ETF holding set. At the quarterly level, the *interconnect* for stock i is the average of the daily *interconnect* in each quarter. We provide more construction details of interconnectedness through ETFs in online Appendix A.

Fig. 2 shows the trend of stock's ETFs interconnectedness and ETF holding number in the sample period. Both the *interconnect* and the ETF holding number for an averaging stock have increased over time. For an averaging stock, it is held by 10 ETFs at the first quarter of 2017 and by nearly 60 ETFs at the second quarter of 2023. The interconnectedness for an averaging stock increased from 1 in 2017 to 2.5 in 2023.

3.2.2. Measuring tail dependence

We construct a dynamic mixture Clayton-survival Gumbel copula to measure the dynamic lower tail dependence. Copula functions and their related extension models are the classic methods for financial risk contagion studies, and the concept of Copula functions was first proposed by Schweizer and Sklar (1983) and has been widely used in the field of financial risk studies since then. As Boccaletti et al. (2006) said, although the CVaR model has overcome the problem of thick-tailed distribution of VaR risk for single factor, but the thick-tailed measure of multi-risk factors seems to be incompetent, so the introduction of the Copula function can be constructed as a joint distribution of the factors and accordingly to achieve the measure of the total systematic risk of the value of the additive. Subsequent studies have made many deformations and extensions to the Copula function to different degrees.

There is a large amount of and still growing body of literature on the copula function due to its flexibility in describing various patterns of dependence structure such as non-linearly, asymmetry, dynamic, and tail dependence (Chabi-Yo et al., 2018; Christoffersen et al., 2012; Hüttner et al., 2020; Sahamkhadam et al., 2022). A copula captures the dependence structure of a multivariate distribution and is defined as a multivariate distribution function with standard uniform margins. There are a variety of copulas in the literature and each copula captures a different dependence structure and dependence degree. Therefore, an appropriate copula function should be selected depending on the nature of financial contagion, which suggests the exploration of extreme dependence (especially in the lower tail) between two markets rather than the widely used correlation in the literature (Wang et al., 2021; Ye et al., 2017). In this paper, we are merely interested in the lower tail dependence and looking for the copula that captures the lower tail dependence feature.

The Clayton copula and Gumbel copula as well as the mixture of the two copulas, with or without time variation, have received much recent attention and have been popularly used in modeling tail dependence (Okimoto, 2008; Wang et al., 2021). As the mixture copula is more flexible and performs better than the single copula (Wang et al., 2021), we construct a dynamic mixture Clayton-

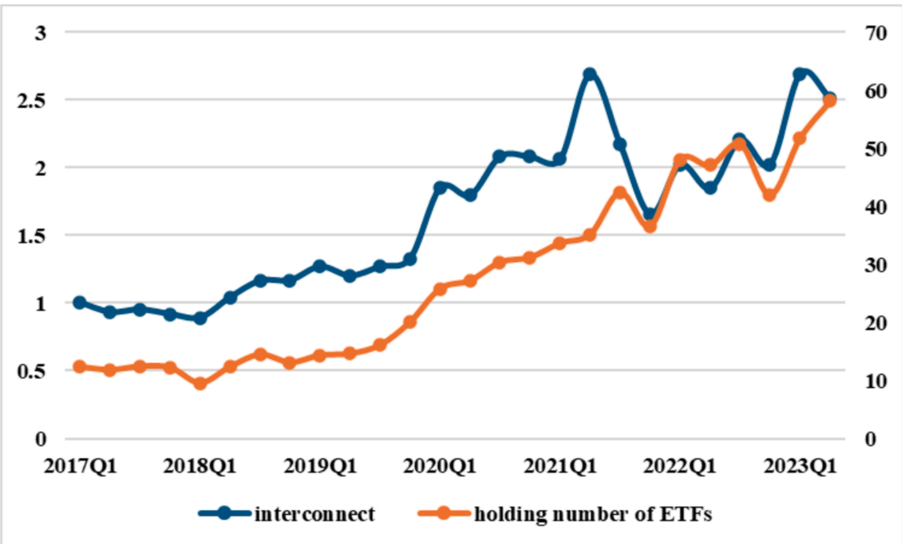


Fig. 2. Interconnectedness and ETF Holding Number of Stocks during Our Sample Period. Notes: this figure presents a graph of the time-series variation in the interconnectedness metric and the number of ETFs held in our calculations over the sample period. Interconnect represents the average degree of Interconnectedness through ETFs calculated according to Equation (1), while the ETF holdings Num indicates how many ETFs hold a single stock on average.

survival Gumbel copula to measure the dynamic lower tail dependence. Since this tail dependence measure is widely used in the literature in financial risk, we will introduce the detailed construction methods and specifications in online Appendix B to save space and for clearer exhibition.

Tail dependence and stock price synchronization are related concepts that measure the co-movement among stock prices, albeit with distinct economic meanings. Specifically, tail dependence focuses on the statistical relationship between extreme events, typically assessed using tools like copulas to capture dependence structures between extreme values of variables. On the other hand, stock price synchronization represents the proportion of market-wide information relative to total information (Kim et al., 2021) and is commonly used as a proxy for information efficiency. Existing studies generally agree that there is a negative relationship between stock price synchronization and market information efficiency (Dasgupta et al., 2010; Gassen et al., 2020), i.e., a higher stock price synchronization implies that less firm-specific information is contained in the stock price, while more noise trading results in poorer market information efficiency. Previous studies have analyzed the factors affecting stock price synchronization from the perspectives such as disclosure (Chan and Hameed, 2006; Xu et al., 2024), financial performance (Magner et al., 2022), and institutional investor governance (Gul et al., 2010; Barka et al., 2023).

Our empirical findings highlight that tail dependence more comprehensively captures stock price co-movement during stress periods. We observe that both tail dependence and stock price synchronization increase when a stock experiences a tail event, indicating a correlation between the two concepts. However, we distinguish the fundamentals driving stock price synchronization from tail dependence. Through our analysis, we discover that interconnectedness via ETFs can account for the portion of tail dependence not explained by stock price synchronization based on fundamentals. This finding implies that ETFs play a role in transmitting non-fundamental risk among stocks.

3.2.3. Stock-level ETFs arbitrage

We define stock-level ETFs arbitrage activity, *Netarbitrage,* as the sum of the netflow pressure of the ETFs that holding the stock. The netflow pressure is motivated by the fact that APs' ETFs arbitrage activity involve the creation and redemption of ETFs shares and will put price pressure on underlying securities (Ben-David et al., 2018; Brown et al., 2021). Specifically, for stock *i* on day *t*, $Netarbitrage_{i,t}$ is defined as

$$Netarbitrage_{i,t} = \frac{\sum_{j=1}^{N} Netflow_{j,t} {}^{*}AMT_{ij,t}}{\sum_{j=1}^{N} AMT_{ij,t}} = \frac{\sum_{j=1}^{N} \left[NAV_{j,t} - NAV_{j,t-1}\left(1+R_{j,t}\right)\right] {}^{*}AMT_{ij,t}}{\sum_{j=1}^{N} AMT_{ij,t}} \tag{2}$$

Where $j \in \{1, \cdots, N\}$ is the set of the ETF holding stock *i*; $NAV_{j,t}$ is the total net asset value of ETF *j* in day *t*, $R_{j,t}$ is the logarithmic return of ETF *j* in day *t*, $AMT_{ij,t}$ is the amount of stock *i* required by ETF *j* when conversion shares, which equals to the required number of shares multiplied by the closing price of the previous trading day.

$Netarbitrage_{i,t}$ sums the netflow of all ETFs that holding stock *i* to measure the overall ETFs arbitrage activity on day *t*. To further illustrate the arbitrage process through creation/redemption of ETFs shares, we distinguish ETFs premium and ETFs discount cases. In

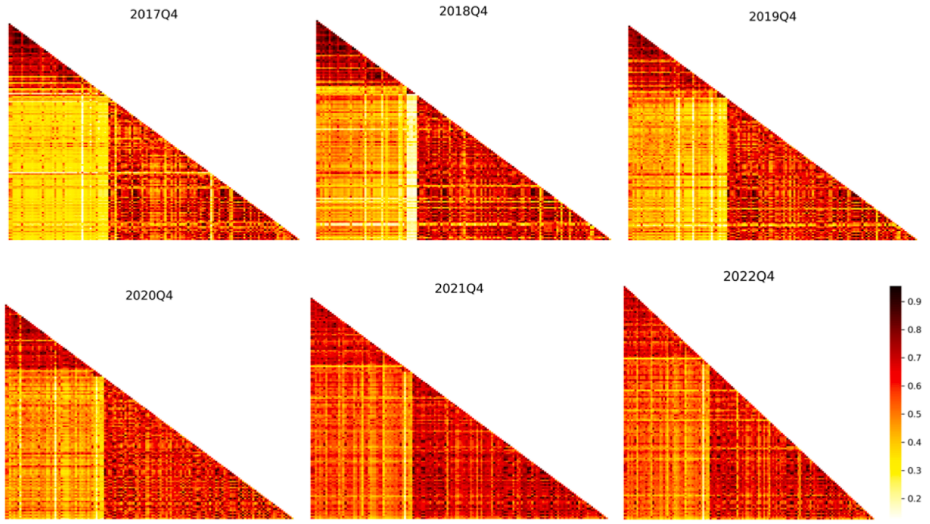
Table 1
Summary Statistics.

	Count	Mean	Sd	Min	P25	P50	P75	Max
Interconnect	64,687	1.3866	1.5331	0.0706	0.3606	0.5467	2.0898	7.1636
Copulas	64,687	0.1848	0.0944	0.0281	0.1156	0.1669	0.2353	0.5671
Instratio	64,687	0.3536	0.2268	0.0011	0.1482	0.3487	0.5337	0.8818
ETF Ratio(%)	64,687	0.4058	0.3657	0.0227	0.1630	0.2895	0.5189	2.5193
Mktvalue	64,687	2.2960	0.9957	0.5689	1.5539	2.1086	2.8653	5.9818
Age	64,687	2.2584	0.8126	0.1187	1.7474	2.3844	2.9844	3.3909
Leverage	64,687	0.4399	0.2032	0.0490	0.2801	0.4316	0.5860	0.9336
ROE	64,687	0.0474	0.0743	−0.5096	0.0129	0.0376	0.0789	0.3797
Stdev(%)	64,687	0.4277	0.1126	0.1609	0.3477	0.4220	0.5022	0.7673
Illiquidity	64,687	0.3596	0.4420	0.0083	0.0973	0.2042	0.4369	3.6229
Netarbitrage(billion)	64,687	0.0411	1.4146	−9.2876	−0.0266	0.0000	0.0375	11.4855
Negarbitrage(billion)	64,687	−0.4461	1.1822	−11.6276	−0.2529	−0.0057	−0.0000	−0.0000
Posarbitrage(billion)	64,687	0.4842	1.3663	0.0000	0.0000	0.0073	0.2506	13.8453

Notes: This table presents the summary statistics for main variables used in this study. $interconnect_{i,T}$ represents the quarterly Interconnectedness with other stocks through ETFs during quarter *T*. $copulas_{i,T}$ represents quarterly tail dependence and computed by averaging stock *i*'s tail dependence with other stocks in quarter *T*. *Instratio* (institutional investor ownership ratio) represents ratio of shares held by institutional investors to total market share. *ETF Ratio* (ETFs ownership ratio) represents ratio of shares held by ETFs to total market share. *Mktvalue* (firm size) represents the logarithm of the total market value. *Age* (firm age) represents logarithm of enterprise age. *Leverage* (firm leverage) represents ratio of total liabilities to total assets. *ROE* (firm performance) represents net income divided by total Equity. *Stdev* (volatility) represents standard deviation of daily returns. *Illiquidity* represents the illiquidity ratio following Amihud (2002). *Netarbitrage* represents the sum of the netflow pressure of the ETFs that holding the stock. $Negarbitrage_{i,t}$ represents the ETFs discount arbitrage activity, which is the sum of the netflow pressure of the ETF holding the stock that experience negative netflow. $Posarbitrage_{i,t}$ represents the ETFs premium arbitrage activity, which is the sum of the netflow pressure of the ETF holding the stock that experience positive netflow.

the case of an ETFs premium, the price of the ETFs exceeds the NAV, APs have an incentive to buy the underlying securities, and ask for newly created ETFs shares. ETFs experience positive netflow under this process. In the case of an ETFs discount, the price of the ETFs is below the NAV, APs have an incentive to redeem ETFs shares and ask for the basket of underlying securities. This generates negative netflow for ETFs. We choose ETFs subsamples according to the sign of their netflow, to measure the ETFs premium and ETFs discount arbitrage activities. $Posarbitrage_{i,t}$ denotes the ETFs premium arbitrage activity, which is the sum of the netflow pressure of the ETF holding the stock that experience positive netflow. $Negarbitrage_{i,t}$ denotes the ETFs discount arbitrage activity, which is the sum of the netflow pressure of the ETF holding the stock that experience negative netflow. At the quarterly level, the ETFs Arbitrage for stock i is the sum of the daily ETFs Arbitrage in the given quarter.

Panel A. Time-Series Heat Map of Interconnectedness of CSI 300 Index Component Stocks.



Panel B. Time-Series Heat Map of Tail Dependence of CSI 300 Index Component Stocks.

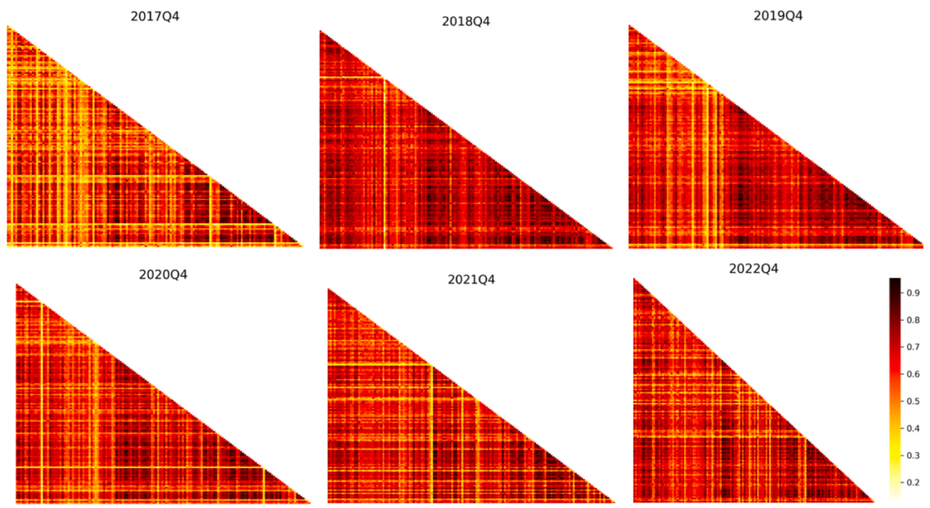


Fig. 3. Time-Series Heat Map of Interconnectedness and Tail Dependence of CSI 300 Index Component Stocks. Notes: this figure illustrates the heat map of Interconnectedness and Tail Dependence of the CSI 300 component stocks at the end of each year during the sample period. Since the index constituents will be adjusted every year, we choose the stocks that belong to the CSI 300 index for the whole sample period as the research object and calculate the interconnectedness and tail dependence matrix between the two stocks and plot the heat map in the graph. Panel A. Time-Series Heat Map of Interconnectedness of CSI 300 Index Component Stocks. Panel B. Time-Series Heat Map of Tail Dependence of CSI 300 Index Component Stocks.

3.2.4. Control variables

We consider several control variables that have been used in prior literature to explain stock tail dependence (Naeem and Karim, 2021). These variables include institutional investor ownership ratio (*Instratio*, ratio of shares held by institutional investors to total market share), ETFs ownership ratio (*ETF Ratio*, ratio of shares held by ETFs to total market share), ETFs vol (a weighted average of the volatility of the ETF holding the stock), firm size (*Mktvalue,* the logarithm of the total market value), firm leverage (*Leverage,* ratio of total liabilities to total assets), firm performance (*ROE,* net income divided by total Equity), firm age (*Age,* logarithm of enterprise age), volatility (*Stdev,* standard deviation of daily returns) and liquidity (*Illiquidity,* the illiquidity ratio follow Amihud (2002)).

3.3. Summary statistics

Table 1 presents the basic statistics of main variables used in this study. Over the sample period, the average interconnectedness is 1.38, with a standard deviation of 1.53, while the average copula is 0.18, with a standard deviation of 0.09. The mean institutional investor holding ratio of the stocks in our sample is 35.36 %, while the mean ETF holding ratio is 0.4058 %. Although ETFs have grown rapidly in recent years, the percentage of their holding is relatively small compared to other institutional investors. However, this smaller percentage of ETF holding has generated substantial arbitrage activity in stocks, with the average net arbitrage flow exceeding 40 million CNY in an averaging quarter, and the discount and premium arbitrage flows being as high as 400 million CNY.

4. Empirical results

4.1. Visualization intuition

Our objective is to test whether ETFs leads to an increase in the tail dependence of underlying securities. To this end, we first exploit the variation in the stock level interconnectedness and tail dependence over time.

Table 2
Baseline Regression.

Dep.Var	*Copulas*					
	(1)	(2)	(3)	(4)	(5)	(6)
Interconnect	0.0246***	0.0274***	0.0263***	0.0193***	0.0198***	0.0209***
	(0.0034)	(0.0031)	(0.0034)	(0.0031)	(0.0031)	(0.0032)
Stdev	0.0508***	0.0416***	0.0394***	0.1046***	0.1056***	0.0991***
	(0.0045)	(0.0043)	(0.0051)	(0.0049)	(0.0049)	(0.0049)
Mktvalue	0.0335***	0.0366***	0.0373***	0.0205***	0.0206***	0.0212***
	(0.0012)	(0.0012)	(0.0013)	(0.0007)	(0.0007)	(0.0007)
Turnover	0.1546***	0.1702***	0.1720***	0.1327***	0.1330***	0.1373***
	(0.0018)	(0.0017)	(0.0019)	(0.0019)	(0.0019)	(0.0019)
Illiquidity	0.0121***	0.0249***	0.0282***	0.0283***	0.0279***	0.0281***
	(0.0017)	(0.0017)	(0.0021)	(0.0012)	(0.0012)	(0.0013)
Instratio		0.1042***	0.1032***	0.0614***	0.0616***	0.0670***
		(0.0031)	(0.0033)	(0.0021)	(0.0021)	(0.0021)
ETFs ownership		−0.1634***	−0.1552***	−0.1170***	−0.1171***	−0.1239***
		(0.0177)	(0.0184)	(0.0181)	(0.0182)	(0.0181)
ETF vol		0.0406***	0.0367***	0.0603***	0.0601***	0.0597***
		(0.0113)	(0.0120)	(0.0128)	(0.0128)	(0.0128)
Leverage		0.0248***	0.0250***	0.0022	0.0027	0.0037
		(0.0048)	(0.0051)	(0.0023)	(0.0023)	(0.0025)
ROE		−0.0391***	−0.0418***	−0.0275***	−0.0280***	−0.0286***
		(0.0047)	(0.0048)	(0.0048)	(0.0048)	(0.0049)
Age		0.0282***	0.0336***	−0.0063***	−0.0062***	−0.0060***
		(0.0023)	(0.0027)	(0.0007)	(0.0007)	(0.0007)
Lag(Y)			0.0093*	0.1252***	0.1233***	0.1098***
			(0.0054)	(0.0071)	(0.0071)	(0.0068)
Constant	0.0084**	−0.1116***	−0.1255***	−0.0040	−0.0049	−0.0059*
	(0.0034)	(0.0066)	(0.0075)	(0.0035)	(0.0035)	(0.0035)
Stock FE	Yes	Yes	Yes	No	No	No
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	No	No	No	Yes	No	No
Province FE	No	No	No	No	Yes	No
Industry Province FE*	No	No	No	No	No	Yes
N	57,120	54,284	45,496	45,723	45,722	45,712
Adjusted R^2	0.7701	0.7961	0.7990	0.7322	0.7328	0.7391

Notes: The table lists the results of the impact of interconnectedness through ETFs on stock tail dependence. The dependent variable *Copulas* measures quarterly tail dependence and computed by averaging stock i's tail dependence with other stocks in quarter T. The independent variable *Interconnect* measures the quarterly interconnectedness with other stocks through ETFs during quarter T. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

For visualization purposes, we take the constituent stocks of the CSI 300 Index as a subsample and snapshot the *interconnect* and *copulas* for the fourth quarter of each year from 2017 to 2022 in Fig. 3. The colors used to represent the degree of interconnectedness (tail dependence) vary from low (yellow) to high (red) in the grids. In Panel A, one observes that the *interconnect* moves to a wide range of red colors from 2017 to 2022, signaling increasing interconnectedness through ETFs between the underlying securities of the CSI 300 Index. The same patten shows in Panel B. In 2017Q4, the *copulas* among the sample stocks are characterized by dispersion. Although some stocks show high tail dependence among each other, there are still many stocks with low tail dependence. With the sharp decline of the China's stock market in 2018, the tail dependence among stocks increases rapidly, and almost all stocks show high tail dependence among each other. As the tension on the stock market has gradually eased, one can see a reduction in tail dependence in 2019. After that, although the Chinese stock market maintains an oscillatory trend, the tail dependence among the sample stocks shows a certain upward trend. Fig. 3 gives us with a simple intuition that the increase in the tail dependence of individual stocks seems to be related to the increase in their interconnectedness through ETFs.

4.2. The effect of interconnectedness through ETFs on tail dependence

Motivated by the heatmap in Fig. 3, we start by looking at whether interconnectedness through ETFs affects the tail dependence of the underlying securities. We use a standard panel regression model to examine the effect of interconnectedness through ETFs on stock tail dependence, and run the following regression:

$$Copulas_{i,T} = \alpha + \beta_1 interconnect_{i,T} + \beta_2 Control_{i,T} + StockFE + TimeFE + \varepsilon_{i,T} \tag{3}$$

Where $copulas_{i,T}$ denotes quarterly tail dependence and computed by averaging stock *i*'s tail dependence with other stocks in quarter *T*, $interconnect_{i,T}$ is the quarterly Interconnectedness with other stocks through ETFs during quarter *T*, $controls_{i,T-1}$ include institutional investor holding ratio, ETF holding ratio, firm size, firm leverage, firm performance, firm age, volatility and Amihud (2002) measure of price liquidity. We also include stock and time fixed effects to control the impact of potentially unobservable factors (e.g., corporate culture and economic cycles). Standard errors are clustered at the stock level.

Table 2 presents the results of regression examining the relationship between interconnectedness through ETFs and tail dependence. In column (1), we run regressions controlling for variables most associated with tail dependence and include indicators of ETFs ownership that are highly correlated with ETF holding. In column (2), we further control for firm fundamental characteristics that potentially affect stock tail dependence. To control for potential serial correlation in tail dependence, we also control for a one-period lag tail dependence in column (3). The coefficients of *Interconnect* are positive and significant at the 1 % level for three model specifications. In other words, we find that interconnectedness through ETFs can amplify the tail dependence among underlying securities. In terms of economic significance, the coefficient of column (3) indicates that a one standard deviation increase in the interconnectedness can lead to tail dependence increasing by 39.95 % of their standard deviation.

In columns (4)-(6) of Table 2, we use alternative fixed effects regressions as a robust test. Specifically, we consider industry-level fixed effect instead of stock-level fixed effects in column (4), to control for unobservable factors such as industry lifecycles and policies. In addition to industry differences, regional differences may also introduce unobservable bias. Recent studies show that differences in risk appetite, industrial policies and business environments across regions may affect the stock tail dependence. To address this bias, we consider a province-level fixed effect instead of an industry-level fixed effect in column (5). Finally, in column (6), we replace the stock fixed effects with industry and province interaction fixed effects to eliminate the potential impact of both industry-level and province-level unobservable bias on tail dependence. The coefficients on the *Interconnect* remain positive and significant across three model specifications, confirming that our results are not driven by potentially unobservable factors.

4.3. Causal inference based on regulatory penalties

The OLS results may be biased due to the endogenous choice. The tail dependence of underlying securities may reduce the participation incentive for authorized participants (APs) with risk averse characteristics. If interconnectedness through ETFs increases the tail dependence of stocks, then ETFs sponsors ex ante require fewer high-risk stocks and more low-risk stocks when setting share conversion rules. As a result, the OLS regressions may underestimate the effect of ETFs on tail dependence. In this section, we use a difference-in-difference approach to address the endogeneity issue.

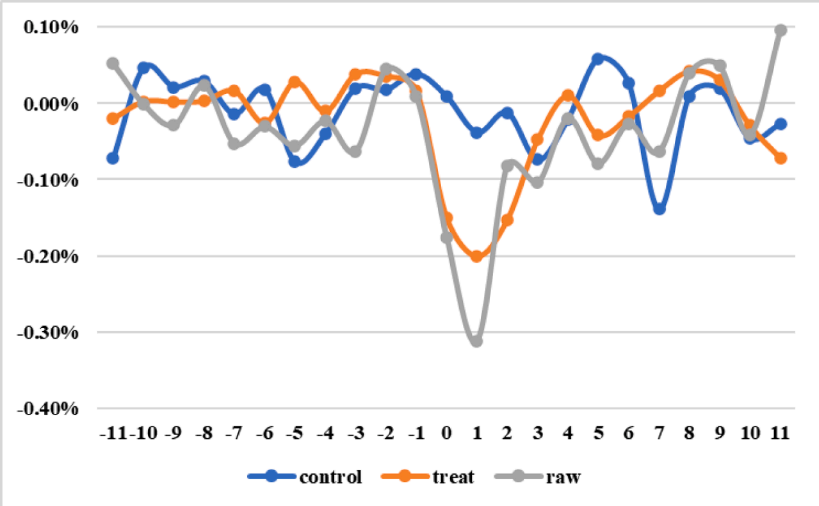
4.3.1. Identification strategy

We use stock penalty information as a quasi-natural experiment for causality identification. Listed firms are penalized for violations by the Shanghai Stock Exchange, the Shenzhen Stock Exchange, the China Securities Regulatory Commission and the Ministry of Finance, and this event has a significant negative impact on the penalized firm's stock price. Although various types of violations are related to the fundamentals of the penalized firms and lead to an endogenous decline in the stock price of the penalized firm, the penalties are not directly related to the fundamentals and stock prices of other listed firms. Based on this economic fact, we use regulatory penalty events as exogenous shocks to firms that have interconnectedness through ETFs with penalized firms, and estimate DID regressions for causal identification using the following model specification:

$$CAR_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t \times Treat_i + \beta_3 Control_{i,t} + \varphi_i + \varepsilon_{i,t} \tag{4}$$

Since we need long-term time series data to calculate the tail dependence between stocks, we cannot construct this indicator at the daily level. To address this issue, we choose the cumulative abnormal return, $CAR_{i,t}$, as the explanatory variable, which is widely used in the study of risk contagion (Campbell et al., 2010). $Post_t$ denotes a time dummy variable, which equals to 1 if the date is after the announcement of the regulatory penalty by the listed company and 0 otherwise, and we use the 10 trading days before and after the announcement of the regulatory penalty as the research window period, so the total window period is 20 trading days. $Treat_i$ represents the treatment and control group dummy variables, and we set the 10 stocks with the highest interconnectedness to the penalized stock as the treatment group stocks, and the 10 stocks with the lowest interconnectedness as the control group stocks. It is worth noting that since we use a staggered DID model, to ensure that the results are not affected by experimental design bias, our treatment group subsample consists of stocks that have never experienced a regulatory penalty event, and the control group subsample consists of stocks that have never experienced regulatory penalties and were never part of the treatment group in the previous period. We also include quarter-level control variables and stock fixed effects in our model.

Panel A. AR for Different Groups of Stocks around the Date of Regulatory Penalty.



Panel B. CAR for Different Groups of Stocks around the Date of Regulatory Penalty.

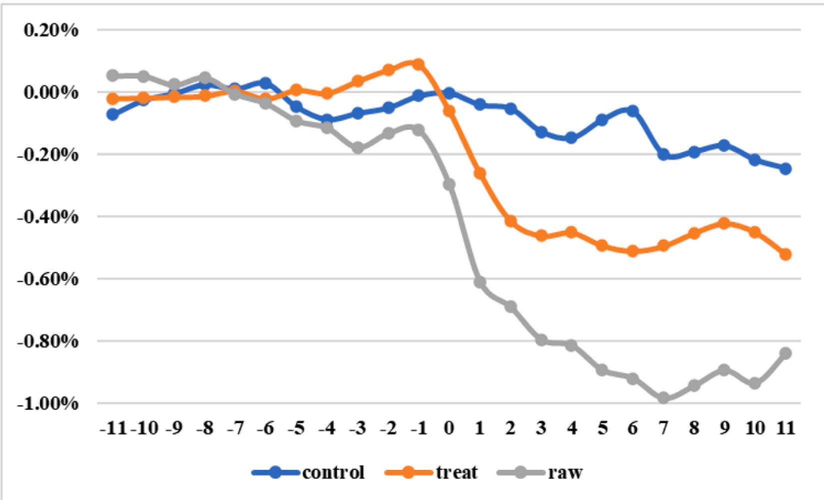


Fig. 4. CAR and AR for Different Groups of Stocks around the Date of Regulatory Penalty. Notes: This figure shows the trend of abnormal returns (AR) and cumulative abnormal returns (CAR) for three types of Difference-in-Differences samples over the sample period. We first use the standard event study approach to analyze whether different levels of interconnectedness lead to penalty risk spillovers. We select penalized stocks, treatment group stocks, and control group stocks, and compute the mean AR and CAR in each period. Then, we plot the AR and CAR for the different sample types during the window period. The horizontal axis represents the relative time window, with negative values indicating the time intervals before the event and positive values indicating the intervals after the event. The vertical axis represents stock returns. Panel A. AR for Different Groups of Stocks around the Date of Regulatory Penalty. Panel B. CAR for Different Groups of Stocks around the Date of Regulatory Penalty.

We first use the standard event study approach to analyze whether different levels of interconnectedness lead to penalty risk spillovers. We select penalized stocks, treatment group stocks, and control group stocks, and compute the mean AR and CAR in each period. Then, we plot the AR and CAR for the different sample types during the window period. As shown in Fig. 4, before the regulatory penalty event, the AR of the three types of stocks do not show significant differences to 0. With the announcement of regulatory penalties for stocks, their AR falls rapidly and does not return to the pre-announcement level until 4 trading days after the announcement, which leads to a rapid decline in the CAR of the penalized stocks after the announcement.

The AR and CAR trends of the penalized stocks are in line with expectations, and we further analyze the treatment and control group stocks to investigate whether ETFs lead to regulatory risk spillovers. The results in Fig. 4 show that the AR and CAR of the treatment group stocks have the same trend as those of the penalized stocks, while the AR and CAR of the control group stocks basically remain at the pre-announcement level. This graphical intuition suggests that interconnectedness may lead to risk contagion from regulatory penalties. in the online Appendix C, we also show the results of T-tests on whether the means are significantly zero in each period for the three samples. The result shows that the CAR of treatment group and penalized stocks are significantly different from zero after the announcement, while the CAR of the control group are significantly zero in each period of the window.

A prerequisite for our quasi-natural experiment in causal identification is that the treatment and control group stocks are not significantly different before the policy. To this end, we plot parallel trend test plots for CAR for the two groups of stocks. As in Fig. 5, the difference between the regression coefficients of CAR for the treatment and control group stocks before the regulatory penalty event is significantly zero, indicating that our identification strategy satisfies the parallel trend hypothesis.

4.3.2. *Empirical results*

After specifying the benchmark regression model at the quarterly level, we first estimate the regression model without any control variables, including fundamental and risk indicator control variables. To reduce the bias introduced by the fixed effects setup, we include only interaction terms in the regressions and implement individual and time fixed effects regression models. In all regression specifications, the coefficients on the interaction term are negative at the 99 % significance level. Compared to stocks with lower interconnectedness to the penalized stocks, stocks with higher interconnectedness to the penalized stocks have significantly lower CAR when a regulatory event occurs.

To better compare the difference in CAR between the two samples before and after the event, we use *POST* as an explanatory variable, and run subsample regressions with the treatment and control groups separately. As in columns (5) and (6) of Table 3, for the control group stocks, there is no significant difference in their CAR before and after the regulatory penalty event. However, for the treatment group stocks, their CARs are significantly lower after the regulatory penalty event. Overall, the results of the quasi-natural causal identification experiment support our conclusion that interconnectedness through ETFs can lead to risk spillovers among stocks.

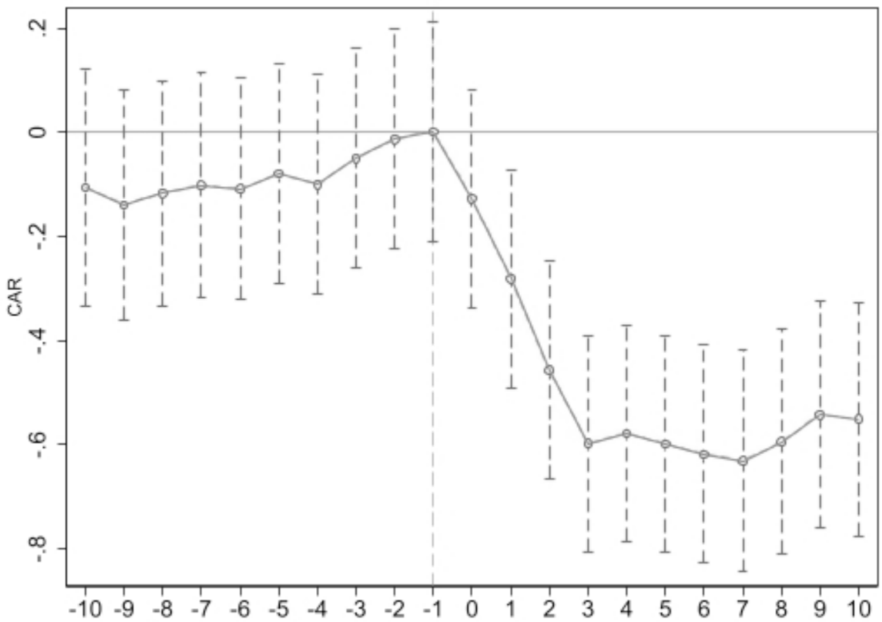


Fig. 5. Parallel Trend Tests Result. otes: This figure presents the results of the parallel trend test for the Difference-in-Differences method. The solid line represents the regression estimates for each period, while the dashed lines indicate the confidence intervals for each estimated coefficient. The horizontal axis represents the relative time window, with negative values indicating the time intervals before the event and positive values indicating the intervals after the event. The vertical axis represents estimated coefficient.

5. Exploring the role of arbitrage activities

As discussed in Section 3, the expansion of ETFs has increased the interconnectedness of underlying securities, leading to higher tail dependence between stocks. Since two stocks connected through ETFs are not necessarily fundamentally related, we need to analyze the underlying channel at the market level. As a highly liquid and low-cost financial innovation, ETFs provide a new trading channel for arbitrageurs in that they can be simultaneously created and redeemed in the primary market and traded in the secondary market. In this section, we examine the role of ETFs arbitrage activity in risk contagion.

5.1. Quarterly evidence

Inspired by the benchmark regression specification, we first examine the evidence of arbitrage activity at the quarterly level. While ETFs arbitrage activity occurs primarily on an intraday basis, the quarterly level analysis aggregates arbitrage activity for each trading day during the quarter and helps to identify those arbitrage activities that are spread across multiple trading days. To this end, we first calculate the intensity of ETFs arbitrage activity for individual stocks at the daily level and aggregate it to the quarterly level and regress it on the *Interconnect*. As shown in the first three columns of Table 4, the *Interconnect* significantly increases net arbitrage activity at the quarterly level, even after controlling for key variables such as the ETFs ownership ratio. For the economic significance of the regression coefficients, a one-standard-deviation increase in *Interconnect* is correlated with an increase of 0.5885 *Netarbitrage*. This amount corresponds to 5.9 % of the standard deviation of *Netarbitrage* in our sample.

In columns (4) to (9) of Table 4, we report the results of the regressions of *Interconnect* on discount and premium arbitrage activity under different model specifications. The regression coefficient of *Interconnect* is significantly negative for discount arbitrage activity. Since discount arbitrage activity itself takes negative values, the significantly negative regression coefficients indicate that interconnectedness through ETFs significantly increase discount arbitrage activity on stocks. In terms of the economic significance of the regression coefficients, a one-standard-deviation increase in the *Interconnect* is correlated with an increase of 0.6076 *Negarbitrage*. This amount corresponds to 7.31 % of the standard deviation of *Negarbitrage* in our sample.

The coefficient of *Interconnect* on premium arbitrage activity, however, is less significant. In regressions that include only market-related control variables, *Interconnect* has a significant positive effect on premium arbitrage activity, but when more stock fundamental control variables are included, the significance of the regression coefficient of *Interconnect* decreases, and when the previous period's premium arbitrage activity is further incorporated into the model to control for the model's potential autocorrelation issue of the

Table 3
DID Regression.

Dep.Var	*CAR*					
	Full Sample				Control Group	Treatment Group
	(1)	(2)	(3)	(4)	(5)	(6)
Post	−0.0094	-0.1394^{**}	-0.1404^{**}		0.0079	-0.4663^{***}
	(0.1231)	(0.0576)	(0.0576)		(0.0662)	(0.0456)
*Post*Treat*	-0.5185^{***}	-0.7019^{***}	-0.7031^{***}	-0.5517^{***}		
	(0.1543)	(0.0730)	(0.0730)	(0.0501)		
Turnover		-3.3335^{**}	-4.7604^{***}	-3.9920^{**}	-13.3161^{***}	-5.0123^{***}
		(1.5568)	(1.5800)	(1.6312)	(4.9030)	(1.5716)
Mktvalue		-0.5441^{***}	-0.6556^{***}	-0.5876^{***}	-0.9131^{***}	-0.6570^{***}
		(0.0885)	(0.0905)	(0.0965)	(0.1463)	(0.1406)
ETFs ownership		−1.1698	−0.1574	3.1148^{**}	−1.2140	−0.9375
		(1.0271)	(1.0401)	(1.4473)	(1.3774)	(2.1634)
ETF vol		1.0917*	0.5258	-2.2701^{***}	0.7930	2.3784*
		(0.5838)	(0.5906)	(0.8213)	(0.7535)	(1.2768)
Illiquidity		−0.0149	−0.0094	−0.0157	0.1265	−0.0196
		(0.0213)	(0.0213)	(0.0216)	(0.1027)	(0.0197)
1/P		-10.7762^{***}	-8.4463^{***}	-5.4068^{**}	16.5521^{***}	-19.6423^{***}
		(2.5632)	(2.6261)	(2.6841)	(5.8231)	(3.0571)
Other Controls	No	No	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	No	Yes	No	No
N	77,861	65,044	65,044	65,044	26,861	38,181
Adjusted R^2	0.2195	0.2307	0.2312	0.2355	0.1224	0.3554

Notes: The table lists the results of DID regressions for causal identification using the following model specification: $CAR_{i,t} = \alpha + \beta_1 Post_t + \beta_2 Post_t \times Treat_i + \beta_3 Control_{i,t} + \varphi_i + \varepsilon_{i,t}$. The independent variable $CAR_{i,t}$ measures the cumulative abnormal return. $Post_t$ denotes a time dummy variable, which equals to 1 if the date is after the announcement of the regulatory penalty by the listed company and 0 otherwise. $Treat_i$ represents the treatment and control group dummy variables. Other control variables include four risk indicators: volatility during the window period (*Stdev*), beta coefficient (*Beta*), idiosyncratic volatility (*Ivol*), and the R2 of the Fama-French three-factor model (*R2*). Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

Table 4
The Impact of Interconnectedness through ETFs on Arbitrage Activities.

Dep.Var	*Netarbitrage*			*Negarbitrage*			*Posarbitrage*		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Interconnect	0.4658***	0.4414***	0.5885***	−0.2376***	−0.5340***	−0.6076***	0.2363***	0.0957	0.0452
	(0.0754)	(0.0792)	(0.0853)	(0.0889)	(0.0872)	(0.0901)	(0.0686)	(0.0674)	(0.0706)
Control	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Other controls	No	Yes	Yes	No	Yes	Yes	No	Yes	Yes
Lag(Y)	No	No	Yes	No	No	Yes	No	No	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	57,035	54,218	45,302	57,035	54,213	45,414	57,037	54,236	45,480
Adjusted R^2	0.1045	0.1134	0.1138	0.2955	0.3353	0.3492	0.2581	0.2876	0.2886

Notes: The table lists the evidence of the impact of interconnectedness through ETFs on arbitrage activities at the quarterly level. The independent variable interconnect measures the quarterly Interconnectedness with other stocks through ETFs during quarter T. In columns 1 to 3, the dependent variable *Netarbitrage measures* the sum of the netflow pressure of the ETFs that holding the stock. In columns 4 to 6, the dependent variable $Negarbitrage_{i,t}$ denotes the sum of the netflow pressure of the ETFs holding the stock that experience negative netflow. In columns 7 to 9, the dependent variable $Posarbitrage_{i,t}$ denotes the sum of the netflow pressure of the ETFs holding the stock that experience positive netflow. The regression model by simple OLS in every three columns, namely columns 1 to 3, columns 4 to 6 and columns 7 to 9, we add different control variables and control for stock and time fixed effects. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

model, the regression coefficient of *Interconnect* is no longer significant. The heterogeneous effect of *Interconnect* on discount and premium arbitrage activity is also consistent with the current study, where ETF holdings affect discount arbitrage activity more than premium arbitrage activity (Ben-David et al., 2018).

5.2. Daily evidence

ETFs provide arbitrageurs with highly liquid and low-cost indexed investment opportunities, and such arbitrage activity tends to occur on an intraday basis. While quarterly regressions provide preliminary evidence on the micro-mechanisms of arbitrage activity, it remains to be tested whether arbitrageurs cause higher tail dependence through intraday arbitrage activity when risk event occurs. based on the causal identification strategy, we discuss whether there is a significant difference in arbitrage activity between individual stocks with different interconnectedness when they experience a regulatory penalty event.

Specifically, we take daily ETFs arbitrage activity of stocks as the explanatory variable and use the DID causality identification specification to test whether interconnectedness through ETFs can increase arbitrage activity. First, we calculate the three types of arbitrage activities of individual ETFs according to equation (2) and calculate the daily ETFs arbitrage activity of stocks according to their ETF holding weights. Then, following the idea of DID quasi-natural experiments, we select the 10 trading days before and after the regulatory penalty event, and use panel regressions after separating the treatment and control groups.

Table 5 presents the results of our daily-level regressions. In columns (1) and (2), we regress daily net ETFs arbitrage activity as the explanatory variable. In column (1), we include both the time dummy variable *Post*, and the intersection *Treat*Post*, while in column (2), we use only the intersection *Treat*Post* and include both time fixed effect and stock fixed effect. Consistent with the quarterly-level

Table 5
Daily Evidence of the Impact of Interconnectedness through ETFs on Arbitrage Activities.

Dep.Var	*Netarbitrage*		*Negarbitrage*		*Posarbitrage*	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	0.0333***		−0.0214*		0.0132	
	(0.0096)		(0.0122)		(0.0200)	
*Post*Treat*	0.0502***	0.0364***	−0.0286***	−0.0181***	0.0025	−0.0329
	(0.0093)	(0.0063)	(0.0084)	(0.0055)	(0.0261)	(0.0213)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	Yes	No	Yes	No	Yes
N	83,885	83,885	75,739	75,739	75,791	75,791
Adjusted R^2	0.0152	0.1171	0.6806	0.8562	0.6995	0.8750

Notes: The table lists the results of DID regressions for causal identification to test the impact of interconnectedness through ETFs on arbitrage activities at the daily level. In columns (1) and (2), we regress daily net ETFs arbitrage activity as the explanatory variable. In column (1), we include both the time dummy variable *Post*, and the intersection *Treat*Post*, while in column (2), we use only the intersection *Treat*Post* and control for both time fixed effect and stock fixed effect. In columns 3 to 6, we further perform quasi-natural experiment causality identification by using discount arbitrage activity and premium arbitrage activity as the dependent variables. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

regression results, we find evidence that ETFs significantly increase the intensity of daily arbitrage activity. Specifically, according to result in column (2), when a regulatory penalty event occurs, stocks with the highest interconnectedness to the penalized stock experience significantly higher arbitrage activity intensity than stocks with the lowest interconnectedness by 3.64 %.

This result reflects the impact of interconnectedness through ETFs on aggregate arbitrage activity at the daily level. As discussed in the previous section, arbitrage traders may engage in discount or premium arbitrage activity depending on the difference between the ETF's market price and the IOPV, but there is heterogeneity in the impact of those activities on the ETF's underlying securities. Therefore, we further perform quasi-natural experiment causality identification by using discount arbitrage activity and premium arbitrage activity as the explanatory variables. In the last four columns of Table 5, interconnectedness with penalized stocks has a significant impact on arbitrage activity. Stocks with higher interconnectedness to penalized stocks experience stronger discount arbitrage activity when a penalty event occurs compared to stocks with lower interconnectedness to the penalized stock.

Meanwhile, as shown in the graphical representation of the parallel trend test of the DID quasi-natural experiment, when a regulatory penalty event occurs in an individual stock, the AR, and the CAR of the penalized stocks as well as the treatment stocks decrease significantly. Guided by this graphical intuition, we investigate whether the cumulative arbitrage activity over the sample period will exhibit the same trend. Specifically, we compute the cumulative arbitrage activity for each trading day during the window period and run a regression with it as the explanatory variable. In Table 6, cumulative arbitrage activity is more significantly affected than daily arbitrage activity. While the regression for daily net arbitrage activity is insignificant, cumulative net arbitrage activity is statistically significant. Stocks with higher interconnectedness to penalized stocks experience stronger arbitrage activity following a regulatory event. Meanwhile, the regressions for cumulative discount arbitrage activity and cumulative premium arbitrage activity are the same as the daily results, with interconnectedness to regulatory penalized stocks leading to stocks experiencing higher discount and premium arbitrage activity.

6. Additional analysis

6.1. Do ETFs provide new risk contagion channel?

Our above empirical evidence shows that ETFs can enhance the connection between their underlying securities. Arbitrage activity between ETFs and their underlying securities can increase the tail dependence of stocks in their baskets. This finding differs significantly from current research on the economic impact of ETF arbitrage activity, and large body of literature find that ETFs themselves provide new source of risk for their underlying securities. For example, in Ben-David et al. (2018), stocks owned by ETFs exhibit significantly higher intraday and daily volatility, and the driving channel appears to be arbitrage activity between ETFs and the underlying stocks. A natural question is whether the enhancing of ETFs on stock tail dependence stems from ETFs themselves, or from the fact that ETFs provide new channel for risk contagion among stocks.

To identify the risk contagion channel of ETFs, we need to isolate the portion of stock tail dependence that originates from ETFs. Specifically, we follow Ben-David et al. (2018) and use stock level ETFs ownership and ETFs volatility to measure the risk pressure originates from ETFs. *ETFs ownership* of stock i at quarter t is defined as the sum across ETFs holding the stock of the dollar value of holdings divided by the stock's capitalization, while *ETFs vol* of stock i at quarter t is defined as a weighted average of the volatility of the ETFs holding the stock. We write the two indicators as follows:

$$ETF\ owernship_{i,t} = \frac{\sum_{j=1}^{J} w_{i,j,t} AUM_{j,t}}{Capitalization_{i,t}}$$

Table 6
The Impact of Interconnectedness through ETFs on Cumulative Arbitrage Activities.

Dep.Var	*Cumulative Netarbitrage*		*Cumulative Negarbitrage*		*Cumulative Posarbitrage*	
	(1)	(2)	(3)	(4)	(5)	(6)
Post	1.1535***		−35.5990***		34.7104***	
	(0.1516)		(0.5274)		(0.5080)	
*Post*Treat*	1.1412***	0.4549***	−8.4978***	−17.8231***	9.9536***	18.5960***
	(0.1957)	(0.1349)	(0.7371)	(0.6215)	(0.7615)	(0.6050)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	Yes	No	Yes	No	Yes
N	79,033	79,033	61,412	61,412	61,252	61,252
Adjusted R^2	0.2401	0.2971	0.4367	0.4286	0.4348	0.4253

Notes: The table lists the results of DID regressions for causal identification to test the impact of interconnectedness through ETFs on cumulative arbitrage activities. In columns 1 and 2, we compute the cumulative arbitrage activity for each trading day during the window period and run a regression with it as the explanatory variable. In column 1, we include both the time dummy variable *Post*, and the intersection *Treat*Post*, while in column 2, we use only the intersection *Treat*Post* and control for both time fixed effect and stock fixed effect. In columns 3 to 6, we estimate the regressions for cumulative discount arbitrage activity and cumulative premium arbitrage activity. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

$$ETF\ Vol_{i,t} = \frac{\sum_{j=1}^{J} w_{i,j,t} AUM_{j,t} Volatility_{j,t}}{Capitalization_{i,t}}$$

where J is the set of ETFs holding stock i; $w_{i,j,t}$ is the weight of the stock in the portfolio of ETF j; $AUM_{j,t}$ is the assets under management of ETF j; $Volatility_{j,t}$ is the quarterly volatility of ETFs j; and $Capitalization_{i,t}$ is the market capitalization of stock i at the end of quarter t.

We then perform a regression of copulas on ETF ownership and ETF volatility, controlling for other potential influences on tail dependence. We use the estimation error as the measure of excess tail dependence (*Excess Copulas*). *Excess Copulas* can be viewed as the portion of stock tail dependence that cannot be explained by ETFs risk. If *Interconnect* still has a significant impact on *Excess Copulas*, we can expect that ETFs provide a new source of risk other than ETFs themselves.

Table 7 presents our results. In columns (1) to (3), we run panel regressions on *Copula* using *ETF Ownership* and *ETF Vol* as explanatory variables. It shows that the coefficients of either *ETF Ownership* and *ETF Vol* are significantly positive, indicating that ETF itself is a source of stock risk. Higher ETF ownership or volatility leads to increased stock tail dependence, consistent with existing literature. Columns (4) to (6) report the results of the regressions of *Interconnect* on *Excess Copula*. Since we use the same control variables as in the baseline regression when calculating excess tail dependence, none of the regression results for the control variables in Columns (4) to (6) are significant, and this suggest that our excess tail dependence metrics satisfy the experimental design. Meanwhile, *Interconnect* has a significantly positive regression coefficient, indicating that interconnectedness through ETFs can explain the portion of tail dependence that cannot be explained by ETFs or other factors, and ETFs provide new channels for risk contagion between stocks.

Another question regarding ETFs and stock tail dependence is whether ETFs enhance non-fundamental tail dependence between underlying securities. The asset allocation of ETFs is based on the holdings of the benchmark index they track, and some indices have a high concentration of holdings, which may lead to strong fundamental connections between underlying securities of ETFs. For example, industry ETFs often hold stocks from a particular industry, and the high tail dependencies between these stocks may stem from their fundamental connections, rather than risk contagion caused by ETF's holding behavior. Thus, we may overestimate the risk contagion role of ETFs. We further investigate whether ETFs lead to non-fundamental tail dependence between underlying securities.

Specifically, we refer to the literature on stock price synchronicity to construct a regression model of stock returns against market index returns or industry index returns (Chan and Hameed, 2006; An and Zhang, 2013; Chan and Chan, 2014). We then use the R^2 of this regression to construct the stock price synchronicity, where synch $= \ln(R^2/(1 - R^2))$. A higher value of *synch* indicates that the stock price is more synchronized. Based on the stock price synchronicity metric, we regress the *Copulas* on synch and other variables that may affect stocks' tail dependence. We then use the regression residuals as a proxy for non-fundamental tail dependence between stocks (*non-fundamental Copulas*) and conduct a panel regression using *Interconnect* as the explanatory variable.

In the regression results presented in Table 8, we first regress *Copulas on* market price synchronicity (MKT *synch*) and industry price synchronicity (IND *synch*) in columns (1) to (3). The results show that the regression coefficients of both MKT *synch* and IND *synch* are significantly positive, indicating that tail dependence among stocks is influenced by market or industry-level fundamental information. On the other hand, columns (4) to (6) report the regression results of *non-fundamental Copulas* on *Interconnect*. Since we use the same control variables as in the baseline regression to account for other potential factors affecting tail dependence, the regression results of the control variables in columns (4) to (6) are not significant, suggesting that our measure of excess tail dependence meets the requirements of the experimental design. The regression coefficient of *Interconnect* on *non-fundamental Copulas* is significantly positive, indicating that interconnectedness through ETFs can explain the portion of tail dependence among stocks that cannot be accounted for by market or industry fundamental information, implying that ETFs transmit non-fundamental risk. We also regress stock-level

Table 7
Do ETFs Provide New Risk Contagion Channel.

Dep.Var	*Copulas*			*Excess Copulas*		
	(1)	(2)	(3)	(4)	(5)	(6)
ETF ownership	0.1138***		0.1536***			
	(0.0095)		(0.0166)			
ETF Vol		0.0569***	0.0294***			
		(0.0061)	(0.0107)			
Interconnect				0.0261***	0.0279***	0.0262***
				(0.0034)	(0.0034)	(0.0034)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	46,187	46,187	46,187	45,816	45,816	45,816
Adjusted R^2	0.8141	0.8135	0.8143	0.0036	0.0040	0.0037

Notes: The table lists the results of regressions to identify the risk contagion channel of ETFs. We perform a regression of copulas on ETF ownership and ETF volatility, controlling for other potential influences on tail dependence. In columns 1 to 3, we run panel regressions on *Copula* using *ETF Ownership* and *ETF Vol* as explanatory variables and control for stock and time fixed effects. In columns 4 to 6, we use the estimation error as the measure of excess tail dependence (*Excess Copulas*) as the dependent variable. Columns 4 to 6 report the results of the regressions of *Interconnect* on *Excess Copula* and control for the stock and time fixed effects. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

Table 8
Do ETFs Drive Non-Fundamental Tail Dependence.

Dep.Var	*Copulas*			*non-fundamental Copulas*		
	(1)	(2)	(3)	(4)	(5)	(6)
MKT synch	−0.0178***		−0.0173***			
	(0.0003)		(0.0003)			
IND synch		−0.3550***	−0.1469***			
		(0.0223)	(0.0214)			
Interconnect				0.0269***	0.0316***	0.0271***
				(0.0031)	(0.0034)	(0.0032)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	46,298	46,298	46,298	45,927	45,927	45,927
Adjusted R^2	0.8224	0.8011	0.8228	0.0039	0.0052	0.0044

Notes: The table lists the results of regressions to show whether ETFs drive non-fundamental tail dependence. We use the R^2 of a regression model of stock returns against market index returns or industry index returns to construct the stock price synchronicity, where synch $= \ln(R^2/(1-R^2))$. In columns 1 to 3, we regress *Copulas on* market price synchronicity (MKT *synch*) and industry price synchronicity (IND *synch*) and other variables. We use the regression residuals as a proxy for non-fundamental tail dependence between stocks (*non-fundamental Copulas*) and conduct a panel regression using *Interconnect* as the dependent variable. Columns (4) to (6) report the regression results of non-fundamental Copulas on Interconnect and control for stock and time fixed effects. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

information efficiency on *Interconnect.* Our empirical findings in online Appendix C highlight that stock price synchronization increases when a stock experiences a tail event.

Finally, we examine the differences between ETFs and mutual funds in enhancing stocks' tail dependence. Although ETFs offer investors convenient trading channels with high liquidity and low costs, they are fundamentally still a type of fund product. Like mutual funds, ETFs construct portfolios by holding a basket of stocks and creating shares. As an important financial product in the capital market, mutual fund also holds many underlying financial assets, leading to tight connections between underlying securities. Unlike mutual funds, however, ETFs can be traded in real-time and redeemed continuously on the secondary market. Therefore, we expect that ETFs exhibit more frequent and individualized characteristics in amplifying stock tail dependencies compared to mutual fund.

To validate our hypothesis, we employ quarterly mutual fund asset holding data and construct a mutual fund connection indicator following the Interconnect method. To assess whether ETFs demonstrate high-frequency variability in risk contagion, we create a market stress dummy variable along with two connection stock stress dummy variables. The market stress dummy variable is set to 1 if the market return for a given quarter falls below −3% and 0 otherwise, using the Shanghai Composite Index as the market benchmark.

Table 9
Comparison of ETFs and Mutual Fund Connection.

Stress dummy	(1) Market Stress	(2)	(3) Individual Stress	(4)	(5) Individual Stress	(6)
Interconnect	0.0289***		0.0308***		0.0267***	
	(0.0035)		(0.0038)		(0.0038)	
stress Interconnect*	0.0088***		0.0025**		0.0082**	
	(0.0034)		(0.0012)		(0.0033)	
fundconnect		0.0174***		0.0186***		0.0174***
		(0.0016)		(0.0017)		(0.0017)
*stress*fundconnect*		0.0029**		0.0006		0.0001
		(0.0012)		(0.0011)		(0.0030)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	No	No	Yes	Yes	Yes	Yes
N	45,927	46,298	44,509	44,877	44,509	44,877
Adjusted R^2	0.7991	0.7991	0.8032	0.8033	0.8037	0.8038

Notes: The table lists the results of comparison of ETFs and mutual fund connection. We create a market stress dummy variable along with two connection stock stress dummy variables. We regress Copulas on ETF and mutual fund connection proxies (*Interconnect* and *fundconnect*) as well as their interactions with the *stress dummy*. In columns 1 to 2, the stress dummy is the market stress dummy variable. It is set to 1 if the market return for a given quarter falls below −3 % and 0 otherwise, using the Shanghai Composite Index as the market benchmark. We control for individual effects but not time effects in the regression analysis. In columns 3 and 4, for the connection stock stress dummy variables, we calculate the quarterly returns of stocks connected to stock I through ETFs. The stress dummy is set to 1 if over 34 % of these connected stocks experience returns below −10 %, and 0 otherwise. In columns 5 and 6, we define a second stress dummy to capture periods of stock stress without market stress. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

We opt for an absolute threshold rather than a percentile due to the absence of major market downturns in our sample period, which might otherwise underestimate market stress.

For the connection stock stress dummy variables, we calculate the quarterly returns of stocks connected to an individual stock through ETFs. The stress dummy is set to 1 if over 34 % of these connected stocks experience returns below −10 %, and 0 otherwise. As market stress periods may coincide with connected stock stress periods, we define a second stress dummy to capture periods of stock stress without market stress. To discern the differences in ETFs and mutual funds in amplifying stock tail dependence, we regress Copulas on ETF and mutual fund connection proxies (Interconnect and fundconnect) as well as their interactions with the stress dummy. Given that the market stress dummy is time-based only, we incorporate individual effects but not time effects when using it in the regression analysis.

As shown in Table 9, we find that the coefficients for *Interconnect* and *fundconnect*, and their interaction with market stress dummy are all significantly positive. Both ETFs and mutual funds increase the tail dependence between stocks, and their impact is more pronounced during market stress periods. However, in columns (3) and (4), the coefficients for *Interconnect* and its interaction with stock stress dummy are significantly positive, while the interaction term for *fundconnect* is insignificant. This suggests that mutual funds react more slowly to risk events due to supervisory restrictions on portfolio adjustments and rebalancing. In contrast, ETFs, with their higher trading frequency and real-time trading capabilities, exacerbate the spread of such short-term risks. In columns (5) and (6), we replace the stock stress dummy, and the results further confirm this view. Overall, we find that compared to mutual funds, ETFs exhibit high-frequency risk contagion characteristics, amplifying the spread of idiosyncratic risks among stocks.

6.2. Robustness checks

In our previous analysis, we controlled for high-dimensional fixed effects and use a quasi-natural experiment based on stock administrative penalties to identify causality, and addressing potential omitted variable and endogeneity issues. In this section, we further conduct a series of robustness checks. Specifically, we perform regressions by replacing the core explanatory variables, constructing instrumental variables for two-stage least squares regression, and using different types of subsamples for regression analysis.

6.2.1. Alternative measurement of interconnectedness through ETFs

As discussed in Section 2.2, the interconnectedness of an individual stock through ETFs is proxied by its average ETF holding similarity with other stocks, and we simultaneously use cosine similarity and the number of holdings intersections between two stocks in the ETF holdings matrix to construct the proxy. To ensure robustness, we also validated our results with several alternative measurements of interconnectedness. To begin with, we only adopt the cosine similarity of stock i's and stock j's ETF holding weight vectors, to measure the ETF holding similarity. Results are reported in column (1) of Table 10. We still find that interconnectedness through ETFs can amplify the tail dependence among underlying securities.

Furthermore, in constructing the interconnectedness proxy in our baseline regression, we chose all ETFs as our sample to build connections between stocks. However, we may ignore the fact that ETFs or stock indexes may construct their constituents based on the potential fundamental correlations between stocks. For example, underlying securities of industry-ETFs are typically from the same

Table 10
Robustness Check: Alternative Variables and IV Estimation.

Method	Alternative Variables			IV Estimation		
	(1)	(2)	(3)	(4)	(5)	(6)
Cosine	0.0025***					
	(0.0003)					
Inter(low fundamental)		0.0018***				
		(0.0002)				
Inter(high institution)			0.0003***			
			(0.0000)			
Interconnect				0.1513***	0.0249***	0.0252***
				(0.0172)	(0.0031)	(0.0031)
Control	Yes	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes	Yes
N	45,885	46,068	45,928	46,149	46,149	46,149
Adjusted R^2	0.7987	0.7987	0.7994	0.7007	0.7270	0.7270

Notes: The table lists the results of robustness checks. We perform regressions by replacing the core explanatory variables and constructing instrumental variables for two-stage least squares regression. In column 1, we adopt the cosine similarity of stock i's and stock j's ETF holding weight vectors to measure the ETF holding similarity. In column 2, we select broad-based ETFs and thematic ETFs as the ETF sample to construct the interconnectedness proxy. In column 3, we select ETFs with institutional investor ownership exceeding 90 % as our sample to construct the interconnectedness proxy. In columns 4 and 5, we conduct regressions using instrumental variables at the industry level and size group level. We compute the mean Interconnect within each industry of stock-quarter group, and we calculate the mean Interconnect for each decile of market capitalization within each quarterly group of firms. We use these mean Interconnect as instruments for stock-level interconnectedness. In column 6, we include both instrumental variables in the regression simultaneously. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

industry, and these stocks often exhibit strong fundamental correlations. To mitigate the potential bias from the fundamental correlations between stocks, we only select broad-based ETFs and thematic ETFs as the ETF sample to construct the interconnectedness proxy. Results are reported in column (2) of Table 10. We still find robust evidence that interconnectedness through ETFs can amplify the tail dependence among underlying securities.

Finally, we find that arbitrage activities serve as an important channel through which ETFs enhance tail dependencies among underlying securities. Due to the high participation of retail investors in the Chinese stock market, both the irrational trading behavior of retail investors and fluctuations of underlying securities prices can influence the ETFs price. This may potentially lead to more arbitrage opportunities in ETFs, making it difficult to accurately identify the impact of arbitrage activity on our results. To ease the impact of retail investors, we only select ETFs with institutional investor ownership exceeding 90 % as our sample to construct the interconnectedness proxy. Results in column (3) of Table 10 show that interconnectedness through ETFs can still amplify the tail dependence among underlying securities.

6.2.2. *Instrumental variable estimation*

As the second robustness check, we instrument firm-level interconnectedness with the interconnectedness of the firms' peer groups in terms of size and sector. We compute the mean *Interconnect* within each industry of stock-quarter group. Likewise, we calculate the mean *Interconnect* for each decile of market capitalization within each quarterly group of firms. We repeat the same examination in our baseline regression. The economic rationale is that combining size and industry determines the interconnectedness through ETFs of the stocks but is uncorrelated with its tail dependence since each industry and size group is relatively large in our database. Accordingly, the exclusion restriction should not be violated. Then, we use these mean *Interconnect* as instruments for stock-level interconnectedness. The regression results are shown in columns (4) to (6) of Table 10. In columns (4) and (5), we conduct regressions using instrumental variables at the industry level and size group level, respectively, while we include both instrumental variables in the regression simultaneously in column (6). The regression results remain consistent with the baseline regression.

6.2.3. *Subsample regression*

The third potential estimation bias arises from sample selection bias. We conduct robustness tests by regressing on specific samples and randomly selected samples. As ETFs track specific indices, market capitalization of a stock is the key factor for inclusion in an index and the ETF's stock pool. We first split our sample into large and small group based on whether their market capitalization exceeds the median for a given quarter and conduct subsample regressions. Moreover, state-owned enterprises are a unique part of the Chinese capital market, and often have a larger chance for being included as underlying securities of certain indices and ETFs. Therefore, we further split our sample into state-owned and non-state-owned enterprises and conduct subsample regressions. Finally, we randomly select half of the stocks from the sample and conduct subsample regressions. The regression results are shown in Table 11. Results are similar to our baseline results: interconnectedness through ETFs can amplify the tail dependence among underlying securities across different sample group.

7. Conclusion

We provide empirical evidence that the arbitrage activity between exchange-traded funds (ETFs) and their underlying securities could propagate non-fundamental shocks from one constituent to the broad cross-section of securities they hold. In other words, ETFs could be a new source of risk contagion.

We first calculate the average of the similarity proxy and copula measurements at the stock level and conduct a series of analyses using a panel of stock-quarterly observations between 2017 and 2023. Our findings indicate that ETFs will increase tail dependence among underlying securities, showing that the average tail dependence of stocks is higher when it has stronger ETF holding similarity with other stocks. After conducting control variables at different level and using high dimensional fixed effect panel regression model specification, the effect remains robust at least in the 99 % confidence interval. To solve the endogeneity problem, we utilize regulatory penalty events for Chinese A-share stocks as quasi-natural experiments. Our results indicate that, in comparison to control group stocks, the cumulative abnormal return for treatment group stocks will significantly decrease by 0.5 % after the regulatory penalty events.

We further investigate the relationship between ETF holdings, arbitrage activity, and tail dependence. Our findings indicate that ETF holding similarity significantly increases stocks' ETF arbitrage activity, and such enhancement is observed primarily in discount arbitrage rather than premium arbitrage, which leads to higher tail dependence among stocks. Daily-level difference-in-difference regression analyses indicate that, in comparison to control group stocks, the ETF arbitrage activity for treatment group stocks will significantly increase after regulatory penalty events. In addition, we conduct a heterogeneity analysis to substantiate our findings. We find that ETFs not only contribute to individual stock risk but also facilitate risk contagion among stocks through their holdings. Meanwhile, interconnectedness through ETFs can explain the portion of tail dependence among stocks that cannot be accounted for by market or industry fundamental information. Both ETFs and mutual funds increase tail dependence between stocks during market stress periods. However, during stock stress periods, only ETFs enhances tail dependence between underlying securities.

There is no doubt that the ETFs structure provides great benefits. Among others, ETFs provide a cheaper and more efficient way for investors to diversify into a broad asset portfolio. At the same time, the results in our paper suggest that they may also introduce non-fundamental risk factors into stock market fluctuations. Non-fundamental risk factors can significantly impact institutional investors who engage in frequent trading. Moreover, these costs may ultimately be passed on to passive individual investors who invest through institutional channels. As investors navigate the financial landscape, careful consideration of such implications becomes essential.

Table 11
Robustness Check: Subsample Regression.

Sample	Small firm (1)	Big firm (2)	SoE (3)	Non-SoE (4)	Random Sample (5)
Interconnect	0.0228***	0.0427***	0.0382***	0.0144**	0.0295***
	(0.0040)	(0.0063)	(0.0039)	(0.0062)	(0.0045)
Control	Yes	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes	Yes
Time FE	Yes	Yes	Yes	Yes	Yes
N	22,620	23,030	33,667	12,260	22,559
Adjusted R^2	0.7913	0.8240	0.7922	0.8191	0.7968

Notes: The table lists the results of robustness checks using different types of subsamples for regression analysis. In columns 1 and 2, we split our sample into large and small group based on whether their market capitalization exceeds the median for a given quarter and conduct subsample regressions. In columns 3 and 4, we split our sample into state-owned and non-state-owned enterprises and conduct subsample regressions. In columns 5 and 6, we randomly select half of the stocks from the sample and conduct subsample regressions. Standard errors are reported in parentheses and clustered at the stock level. * denotes significance at the 10 % level, ** denotes significance at the 5 % level, and *** denotes significance at the 1 % level.

Acknowledgement

Jiang acknowledges support from the National Natural Science Foundation of China (NSFC) (No.72072193, 71872195, 72342019) and National Social Science Fund of China (No.22&ZD063). Ning acknowledges support from the National Natural Science Foundation of China (NSFC) (No.72403203) and Humanities and Social Science Research of Ministry of Education of China (No. 24YJC790138). The authors contributed equally to this paper. All errors and omissions are our own responsibility. The authors have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jimonfin.2024.103194.

References

Amihud, Y., 2002. Illiquidity and stock returns: Cross-section and time-series effects. J. Financ. Mark. 5 (1), 31–56.
An, H., Zhang, T., 2013. Stock price synchronicity, crash risk, and institutional investors. J. Corp. Financ. 21, 1–15.
Anadu, K., Kruttli, M., Mccabe, P., Osambela, E., 2020. The shift from active to passive investing: Risks to financial stability? Financ. Anal. J. 76 (4), 23–39.
Antoniou, C., Li, F.W., Liu, X., Subrahmanyam, A., Sun, C., 2023. Exchange-traded funds and real investment. Rev. Finan. Stud. 36 (3), 1043–1093.
Barka, Z., Benkraiem, R., Hamza, T., Lakhal, F., Vigne, S., 2023. Institutional investor horizon and stock price synchronicity: Do product market competition and analyst coverage matter? Int. Rev. Finan. Anal. 89, 102733.
Barucca, P., Mahmood, T., Silvestri, L., 2021. Common asset holdings and systemic vulnerability across multiple types of financial institution. J. Financ. Stab. 52, 100810.
Ben-David, I., Franzoni, F., Kim, B., Moussawi, R., 2023. Competition for attention in the ETF space. Rev. Finan. Stud. 36 (3), 987–1042.
Ben-David, I., Franzoni, F., Moussawi, R., 2018. Do ETFs increase volatility? J. Financ. 73 (6), 2471–2535.
Bhojraj, S., Mohanram, P., Zhang, S., 2020. ETFs and information transfer across firms. J. Account. Econ. 70 (2), 101336.
Boccaletti, S., Latora, V., Moreno, Y., Chavez, M., Hwang, D.-U., 2006. Complex networks: Structure and dynamics. Phys. Rep. 424 (4), 175–308.
Box, T., Davis, R.L., Fuller, K.P., 2019. ETF competition and market quality. Financ. Manag. 48 (3), 873–916.
Brown, D.C., Davies, S.W., Ringgenberg, M.C., 2021. ETF arbitrage, non-fundamental demand, and return predictability. Rev. Financ. 25 (4), 937–972.
Campbell, C.J., Cowan, A.R., Salotti, V., 2010. Multi-country event-study methods. J. Bank. Financ. 34 (12), 3078–3090.
Capponi, A., Glasserman, P., Weber, M., 2020. Swing pricing for mutual funds: Breaking the feedback loop between fire sales and fund redemptions. Manag. Sci. 66 (8), 3581–3602.
Chabi-Yo, F., Ruenzi, S., Weigert, F., 2018. Crash sensitivity and the cross section of expected stock returns. J. Financ. Quant. Anal. 53 (3), 1059–1100.
Chan, K., Chan, Y.-C., 2014. Price informativeness and stock return synchronicity: Evidence from the pricing of seasoned equity offerings. J. Financ. Econ. 114 (1), 36–53.
Chan, K., Hameed, A., 2006. Stock price synchronicity and analyst coverage in emerging markets. J. Financ. Econ. 80 (1), 115–147.
Chernenko, S., Sunderam, A., 2020. Do fire sales create externalities? J. Financ. Econ. 135 (3), 602–628.
Choi, J., Hoseinzade, S., Shin, S.S., Tehranian, H., 2020. Corporate bond mutual funds and asset fire sales. J. Finan. Econ. 138 (2), 432–457.
Christoffersen, P., Errunza, V., Jacobs, K., Langlois, H., 2012. Is the potential for international diversification disappearing? A dynamic copula approach. Rev. Finan. Stud. 25 (12), 3711–3751.
Cont, R., Schaanning, E., 2019. Monitoring indirect contagion. J. Bank. Financ. 104, 85–102.
Coval, J., Stafford, E., 2007. Asset fire sales (and Purchases) in equity markets. J. Financ. Econ. 86 (2), 479–512.
Da, Z., Shive, S., 2018. Exchange traded funds and asset return correlations. Eur. Financ. Manag. 24 (1), 136–168.
Dasgupta, S., Gan, J., Gao, N., 2010. Transparency, price informativeness, and stock return synchronicity: Theory and evidence. J. Financ. Quant. Anal. 45 (5), 1189–1220.
Duarte, F., Eisenbach, T.M., 2021. Fire-sale spillovers and systemic risk. J. Financ. 76 (3), 1251–1294.
Egger, P.H., Li, J., Zhu, J., 2023. The network and own effects of global-systemically-important-bank designations. J. Int. Money Financ. 136, 102879.
Fulkerson, J.A., Jordan, S.D., Travis, D.H., 2022. ETF Arbitrage and daily cash flow. J. Invest. 31 (3), 98–121.
Gassen, J., Skaife, H.A., Veenman, D., 2020. Illiquidity and the measurement of stock price synchronicity. Contemp. Account. Res. 37 (1), 419–456.
Girardi, G., Hanley, K.W., Nikolova, S., Pelizzon, L., Sherman, M.G., 2021. Portfolio similarity and asset liquidation in the insurance industry. J. Finan. Econ. 142 (1), 69–96.

Glasserman, P., Young, H.P., 2016. Contagion in financial networks. J. Econ. Lit. 54 (3), 779–831.
Glosten, L., Nallareddy, S., Zou, Y., 2021. ETF activity and informational efficiency of underlying securities. Manag. Sci. 67 (1), 22–47.
Gul, F.A., Kim, J.-B., Qiu, A.A., 2010. Ownership concentration, foreign shareholding, audit quality, and stock price synchronicity: Evidence from China. J. Financ. Econ. 95 (3), 425–442.
Guo, M., Li, Z., Tu, Z., 2012. A unique "T+1 Trading Rule" in China: Theory and evidence. J. Bank. Financ. 36 (2), 575–583.
Hanley, K.W., Hoberg, G., 2012. Litigation risk, strategic disclosure and the underpricing of initial public offerings. J. Financ. Econ. 103 (2), 235–254.
Hautsch, N., Schaumburg, J., Schienle, M., 2015. Financial network systemic risk contributions. Eur. Finan. Rev. 19 (2), 685–738.
He, W., Luo, J.-H., 2018. Agency problems in firms with an even number of directors: Evidence from China. J. Bank. Financ. 93, 139–150.
Hsiao, C.-L., Chiu, Y.-B., 2024. Financial contagion and networks among the oil and BRICS stock markets during seven episodes of crisis events. J. Int. Money Financ. 144, 103081.
Huang, D., Jiang, F., Tu, J., Zhou, G., 2015. Investor sentiment aligned: A powerful predictor of stock returns. Rev. Financ. Stud. 28 (3), 791–837.
Huang, S., O'Hara, M., Zhong, Z., 2021. Innovation and informed trading: Evidence from industry ETFs. Rev. Financ. Stud. 34 (3), 1280–1316.
Hüttner, A., Scherer, M., Gräler, B., 2020. Geostatistical modeling of dependent credit spreads: Estimation of large covariance matrices and imputation of missing data. J. Bank. Financ. 118, 105897.
Israeli, D., Lee, C.M.C., Sridharan, S.A., 2017. Is there a dark side to exchange traded funds? An information perspective. Rev. Acc. Stud. 22 (3), 1048–1083.
Jiang, F., Lee, J., Martin, X., Zhou, G., 2019. Manager sentiment and stock returns. J. Financ. Econ. 132 (1), 126–149.
Jones, C.M., Shi, D., Zhang, X., Zhang, X., 2023. Retail trading and return predictability in China. J. Financ. Quant. Anal. 1–37.
Kelley, E.K., Tetlock, P.C., 2017. Retail short selling and stock prices. Rev. Financ. Stud. 30 (3), 801–834.
Kim, Y., Su, L.N., Wang, Z., Wu, H., 2021. The effect of trade secrets law on stock price synchronicity: Evidence from the inevitable disclosure doctrine. Acc. Rev. 96 (1), 325–348.
Liu, R., Si, H., Miao, M., 2022. One False Step Can Make a Great Difference: Does Corporate Litigation Cause the Exit of the Controlling Shareholder? J. Corp. Financ. 73, 102192.
Magner, N., Lavín, J.F., Valle, M.A., 2022. Modeling synchronization risk among sustainable exchange trade funds: A statistical and network analysis approach. Mathematics 10 (19), 3598.
Menkveld, A.J., Yueshen, B.Z., 2019. The flash crash: A cautionary tale about highly fragmented markets. Manag. Sci. 65 (10), 4470–4488.
Naeem, M.A., Karim, S., 2021. Tail Dependence between Bitcoin and Green Financial Assets. Econ. Lett. 208, 110068.
Okimoto, T., 2008. New evidence of asymmetric dependence structures in international equity markets. J. Financ. Quant. Anal. 43 (3), 787–815.
Raddant, M., Kenett, D.Y., 2021. Interconnectedness in the global financial market. J. Int. Money Financ. 110, 102280.
Sahamkhadam, M., Stephan, A., Östermark, R., 2022. Copula-based Black-Litterman portfolio optimization. Eur. J. Oper. Res. 297 (3), 1055–1070.
Pan, K., Zeng, Y., 2017. ETF Arbitrage under Liquidity Mismatch. Harvard University. Unpublished working paper.
Sommers, F., 2024. Federal reserve system exchange-traded funds holdings: An operational observation. OpsRisk Limited. Unpublished Working Paper.
Schweizer, B., Sklar, A., 1983. Probabilistic Metric Spaces. Probab. Theory Relat. Field. 26 (3).
Tan, L., Zhang, X., Zhang, X., 2023. Retail and institutional investor trading behaviors: Evidence from china. Annu. Rev. Financ. Econ 16.
Wang, H., Yuan, Y., Wang, T., 2021. The dynamics of cross-boundary fire—Financial contagion between the oil and stock markets. J. Futur. Mark. 41 (10), 1655–1673.
Wu, W., Zhu, F., 2023. ETF ownership and informational efficiency of underlying stocks: Evidence from China. Pac. Basin Financ. J. 79, 102005.
Xu, D., Huang, J., Ren, X., Ye, M., 2024. ESG report textual similarity and stock price synchronicity: Evidence from china. Pac. Basin Financ. J. 85, 102343.
Xu, L., Yin, X., 2017. Does ETF trading affect the efficiency of the underlying index? Int. Rev. Financ. Anal. 51, 82–101.
Ye, W., Luo, K., Liu, X., 2017. Time-varying quantile association regression model with applications to financial contagion and var. Eur. J. Oper. Res. 256 (3), 1015–1028.