Quantitative Finance

ISSN: (Print) (Online) Journal homepage: www.tandfonline.com/journals/rquf20

Tail risk aversion and backwardation of index futures

Jufang Liang, Dan Yang & Qian Han

To cite this article: Jufang Liang, Dan Yang & Qian Han (2024) Tail risk aversion and backwardation of index futures, Quantitative Finance, 24:3-4, 387-407, DOI: 10.1080/14697688.2024.2330612

To link to this article: https://doi.org/10.1080/14697688.2024.2330612

Quantitative Finance, 2024
Vol. 24, Nos. 3–4, 387–407, https://doi.org/10.1080/14697688.2024.2330612

Tail risk aversion and backwardation of index futures

JUFANG LIANG †, DAN YANG† and QIAN HAN‡*

†School of Finance and Statistics, Hunan University, 109 Shijiachong Road, Yuelu District, Changsha, Hunan 410079, People's Republic of China
‡Lingnan College, Sun Yat-Sen University, 135 Xingang Xi Road, Haizhu District, Guangzhou, Guangdong 510275, People's Republic of China

(*Received 26 January 2023; accepted 7 March 2024; published online 12 April 2024*)

We show that tail risk aversion, proxied by the skewness risk premium implied from the SSE 50 ETF options market, explains a significant proportion of the unusually deep backwardation of index futures during the 2015 Chinese stock market crash, while traditional factors such as non-synchronous trading, spot return, volatility and liquidity, all fail to explain the backwardation. These empirical results imply that investors' concern over the crash risk causes speculators to charge a high 'insurance premium' on hedgers. On the other hand, short-sale constraints, namely high security borrowing costs and other barriers, prevent reverse arbitrage such that the deep backwardation of index futures persists.

Keywords: Futures basis; Skewness risk premium; Tail risk aversion

JEL Classifications: G13, G14

'The opportunity to hedge is a valuable service furnished by speculators in the expectation of a financial return and that return is indeed earned. Stated in this fashion, it is no different from the proposition that brokers or barbers expect to earn a living from their occupations and usually do.'

'Returns to Speculators: Telser versus Keynes,' Paul H. Cootner, Journal of Political Economy, 1960.

1. Introduction

China Financial Futures Exchange (CFFEX) launched the CSI 300 index futures contracts (IF) on 16th April 2010, and 5 years later, the CSI 500 and SSE 50 index futures contracts (IC and IH respectively). During the 2015 market crash, the trading volume of the stock index futures markets increased rapidly, while the IC contracts repeatedly hit the daily lower price limit of -10%.

More interestingly, from 25th June, all three index futures contracts with various maturities manifested sustained deep backwardation. On 2nd September 2015, the monthly bases of all nearest-maturity contracts, as represented by the logarithmic difference between the futures and the spot prices, were -15.08%, -12.72% and -10.48% for IC, IF and IH respectively. From another perspective, the average bases from 26th June to 30th September 2015 reached -5.24%, -3.90% and -2.71%, which were never observed since launching the contracts. These extraordinary negative bases raised regulators' and investors' wide concerns, as even an average basis of -3% would imply an annualized hedging cost of 36% by roll-over!

Index futures were accused of being responsible for the market crash by supervisory authorities and some medias because of their deep backwardation and extremely high trade volume. Under heavy pressure from both regulators and the public, the CFFEX announced on 25th August 2015 that starting 26th August, two measures would be adopted to curb speculative trading in the index futures markets. First, any single day's total opening position greater than 600 contracts would be considered abnormal trade and be subject to increased scrutiny. Second, the clearing fees for intra-day trades would be adjusted upward to 1.15 basis points.

With the crisis developing further and the effect of government bailout measures quickly diminishing, 2nd September witnessed the CFFEX announcing yet another round of measures to curb speculative trade in the futures markets. Starting from 7th September, on any single day non-hedging trading

*Corresponding author. Email: hanq37@mail.sysu.edu.cn

of over 10 contracts would be considered abnormal. Moreover, the clearing fees for intra-day trades would be adjusted dramatically from 1.15 to 23 basis points. With these two rounds of measures, index futures trading in China nearly came to a complete stop.

Despite these efforts, the negative bases of index futures persist into the end of 2015. These deep and prolonged negative bases in all index futures contracts are very rare in the global futures markets, as usually arbitrageurs would kick in and profit from the negative basis. In this paper, we aim to understand the causes of this unique phenomenon.

The literature has documented many factors which can result in changes of index futures basis. The classic cost-of-carry model suggests a negative basis when the expected dividend yield goes beyond the risk-free rate. Harris (1989), Antoniou and Garrett (1993) attribute the index price distortions to non-synchronous trading. Chen *et al.* (1995) argue that the return volatility of spot is an important factor for the negative basis during the U.S. 1987 market crash. Roll *et al.* (2007) show that the liquidity of spot and futures markets affects the convenience of arbitrage which in turn affects the basis dynamics.

Our first step is to test these well-documented factors for their explanatory power on the Chinese futures markets. We construct a method to imply the fair values of index futures basis when non-synchronous trading of index-constituent stocks such as hitting price limits and trade suspensions may result in the distortion of indices. We also conduct various regression analyses with spot volatility and market liquidity as explanatory variables. To our surprise, none of these factors is significant in explaining the deep backwardation observed in the Chinese index futures markets.

We then turn to the hedging pressure hypothesis proposed in the derivatives literature. Keynes (1930) and Hicks (1946) argue that in the commodity futures market, speculators risk themselves by entering the market to trade, and hedgers are mainly spot traders who sell futures to hedge. Therefore, speculators are willing to continue to participate in transactions only when the futures price is maintained at the price they expect. When the tail risk aversion is sufficiently high, the hedging pressure increases and speculators face greater risk. At this point, speculators will be less willing to hold long positions as opponents, therefore place lower bids for futures contracts, thus lowering the futures bases.

To test this hypothesis, we follow Feunou *et al.* (2018) to construct the skewness risk premium (hereinafter SRP),† a proxy of market tail risk aversion and hence hedging pressure, based on the SSE 50 ETF option price panel. Our Vector-Autoregression (VAR) analysis and Granger causality tests show that a significant proportion of the negative bases of the index futures in our sample can be explained by the tail risk aversion triggered by market turmoil. Impulse response analysis and variance decomposition of the prediction errors further find that about 29% to 44%, 26% to 57% and 21% to 35% forecasting error variances can be attributed to SRP for IC, IF and IH contracts, respectively. These results are robust to variable orders in the VAR model and a different sample covering the COVID-19 virus spread. Substituting iVIX‡ for the SRP suggests that the iVIX has much weaker explanatory power to the negative bases than the SRP.

To test whether those effects exist in general or just during the market turmoil, we use a multivariate version of the multi-quantile Conditional Autoregressive Value at Risk (hereinafter referred to as MVMQ-CAViaR) and extends our sample to estimate the tail dependence between SRP and basis in the long run. Results derived from the MVMQ-CAViaR model indicate that there is significant tail dependence between the index futures bases and SRP, and this dependency does not exist at the center of endogenous variables distribution (i.e. at the 0.5 quantile). Overall, our results suggest that an increase of tail risk premium can result in index futures backwardation only when tail events occur.

This paper fits the literature on the role of hedging pressure on derivatives pricing in an incomplete market. In the classic cost-of-carry model, the assumption of complete market implies a stable relationship between futures and spot price. Even when there exist trading costs, the basis should stay within a narrow no-arbitrage bound. Our empirical results, however, prove that if the market is incomplete, the cost-of-carry model would be a failure for pricing index futures, and more specifically, the pervasive tail risk premium, under short sale constraints in the spot market, would result in the failure of the law of one price, and it becomes a significant factor in index futures price determination.

Our findings carry practical implications. Chinese regulators and investors need to understand the true reason of why deep backwardation of index futures occurs during the market crash. The deep backwardation of CFFEX index futures made the contracts being accused of being responsible for the market crash and the exchange had to take the sternest policy in history to restrict futures trading, which almost made the index futures markets dysfunctional. Our Granger causality statistics show that the hedging pressure induced by tail risk aversion, under short sale constraints in spot markets, is a significant driver of the futures backwardation. Results from MVMQ-CAViaR model also indicate that there is significant tail dependence between index futures backwardation and tail risk aversion. Therefore, reasonable policies after the market crash should focus on reducing market frictions, i.e. enriching the basket of hedging tools by launching new financial derivatives and relaxing short sale constraints, rather than the continuation of trading restrictions.

The remainder of this paper is structured as follows. Section 2 reviews the literature and develops testable hypotheses. Section 3 introduces the empirical methodology. Section 4 presents empirical results. Section 5 provides further discussions. Section 6 concludes.

† Besides being a proxy of tail risk premium, the SRP is also used as proxy of uncertainty, e.g. Feunou *et al.* (2018) or fears, e.g. Bollerslev and Todorov (2011). In both studies, tail risk premium, uncertainty and fears have the same implication. In this paper, the term of tail risk premium is used.

‡ iVIX is a model-free SSE 50 ETF option implied volatility index which is similar to the CBOE VIX. It has not been published since 22nd February 2018.

2. Literature review and hypotheses development

The cost-of-carry model implies that in a complete market negative basis only appears in times when the risk-free interest rate is far below the expected dividend yield. However, this is clearly not the case for the Chinese market in the period from June to December 2015 as the dividend payments are typically concentrated between May and July. Moreover, the dividend rate of Chinese A-shares has long remained at an obviously too low level to justify the deep backwardation.

MacKinlay and Ramaswamy (1988), Harris (1989), and Antoniou and Garrett (1993) have noted that non-synchronous trading had a major impact on the basis distortion during the 'Black Monday' of 1987. This seems to be a valid hypothesis to test for our case. During the 2015 Chinese market crash, many A-share listed companies voluntarily requested trading suspensions to avoid further price drops. In addition, there were a large proportion of index-constituent stocks that hit the daily upper or lower price limit (positive and negative 10%, respectively) during the market crash. On 24th August 2015, for example, 2186 out of roughly 3400 stocks in the A-share market closed at the daily lower limit. In calculating the index, the price of a suspended stock on the last trading day before the suspension and the price at the upper and lower limits are used. Therefore, it is reasonable to deduce that the presence of suspension and price limit may cause the distortion of index, hence a distorted basis.

We propose to recover the 'true' index level using information from the prices of ETF, the total return index, and the net return index, the markets of which were all in continuous operation during the crash period. However, we find that only on a handful of trading days, the severe negativity of the index futures bases can be explained by non-synchronous trading of individual stocks.

Chen *et al.* (1995) maintain that volatility could play a significant role in changes of basis when investors adopt 'portfolio insurance' strategy, i.e. investors invest in a tailored portfolio and use index futures to hedge risks to reap excess returns. When the market volatility rises, on average investors tend to increase their short positions in index futures. Therefore, the higher the volatility is, the lower the basis. Chen *et al.* (1995) confirm their theoretical predictions through an empirical investigation. Hemler and Longstaff (1991) provide consistent results with this volatility hypothesis. However, we are highly suspicious that volatility is sufficient to explain the futures discount in the Chinese case, because although the market gradually stabilized after 15th September 2015, the negativity of bases still lingered. It is clear that a rigorous re-examination for the Chinese market is warranted.

Roll *et al.* (2007) discuss the relationship between the spot market liquidity and the corresponding index futures pricing efficiency. Empirically, they find a positive correlation between the mean reversion speed of basis and market liquidity. The effect of liquidity is more profound for inactive contracts with longer time to maturity. Han and Pan (2017) find that since July 2015, regulators in China restricted transactions on both the futures and stock markets resulted in shift in the arbitrage boundary and the breakdown of the two-way causality relation between liquidity and the absolute futures-cash basis. For comparison, we also consider liquidity in our following empirical tests.

In another strand of literature, Hirshleifer (1989), Bessembinder (1992) and De Roon *et al.* (2000) find that futures risk premia may be associated with hedging pressure. A study similar to ours, Acharya *et al.* (2013), shows that the risk premium of commodity futures associated with producer hedging demand rises when speculative activity reduces and limits to financial arbitrage generate limits to hedging by producers, therefore affecting equilibrium commodity supply and prices.

The options pricing literature has also documented close relationships between hedging demand and option prices. Bollen and Whaley (2004), for example, report that changes in implied volatility are directly related to net buying pressure from public order flow. And interestingly, changes in S&P 500 index option implied volatility are most strongly affected by buying pressure of index puts, while those of individual stock option implied volatility are dominated by calls. Garleanu *et al.* (2008) model demand-pressure effect on option prices and show empirically that hedging demand helps to explain the overall expensiveness and skew pattern of index option prices.

Measuring hedging pressure becomes an important research topic in the recent literature. Given the unobservable nature of hedging pressure and lack of individual trading records, many scholars suggest to measure hedging pressure through rare events and the associated risk premium. Option pricing models proposed in Liu *et al.* (2005), Bates (2008), Drechsler and Yaron (2011), and Drechsler (2013) have indicated that rare events may have significant impact on asset prices. Liu *et al.* (2005) argue that price jumps represent ambiguity or uncertainty because they have rarely been present in historical data. Given the well-known difficulty of performing statistical analyses and draw inferences with limited data in identifying price jumps, Bollerslev and Todorov (2011), Feunou *et al.* (2018) propose using option prices and high-frequency futures prices instead to estimate the difference between the downside variance risk premium and the upside variance risk premium, namely Skewness Risk Premium (SRP), as a measurement of the tail risk premium which is induced by rare events and market uncertainty.

In this paper, we choose to use skewness risk premium implied from option prices as a proxy of hedging pressure to analyze the interaction between hedging pressure and the basis of index futures. We find that the perceived stock market tail risk increases the hedging pressure on, hence increasing the risk premium of, index futures. To our best knowledge, this is the first empirical study examining the effect of tail events on the futures basis when short sale constraints exist.

3. Methodology

3.1. The futures backwardation

As argued above, during the 2015 market crash, the Chinese stock market index prices could be distorted due to frequent suspensions and/or hitting daily price limits of individual index-constituent stocks, i.e. the so-called non-synchronous trading problem. Non-synchronous trading may cause a misleading futures basis as stale spot prices will be

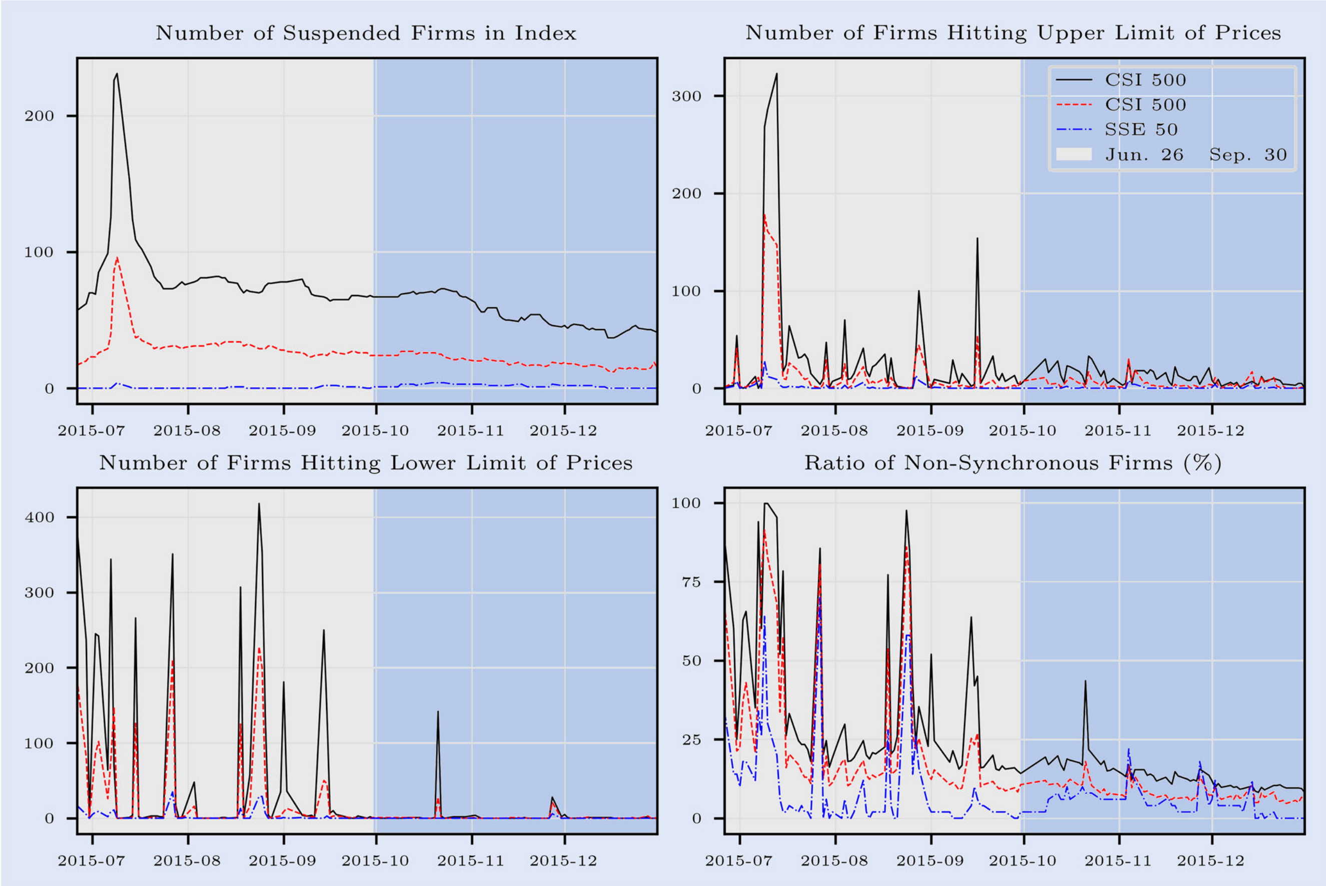

Figure 1. Numbers and ratio of non-synchronous trading firms of index-component stocks.
The greyed regions correspond to the period from 26th June 2015 to 30th September 2015.

incorporated in updating the index level. Figure 1 demonstrates how serious this problem may be during our sample. The upper left subfigure reports the number of suspended index-constituent stocks from 26th June to 31st December 2015. On average, there are 85 and 32 index-constituent stocks which have been suspended from trading in the CSI 500 and CSI 300 indices, respectively. On 9th July, the number of suspended stocks peaked at 231 and 96. The upper right subfigure reports the number of stocks hitting the upper price limit on each date in the subsample. The average number is 33 and 16, and the maximum occurs on 9th July, with 323 and 147 stocks hitting the upper limit, for the CSI 500 and CSI 300 indices respectively. The lower left subfigure reports the number of stocks hitting the lower price limit on each date. Its peaks, 353 and 228, arrive on 25th August and 24th August for the CSI 500 and the CSI 300 indices, respectively. The last subfigure presents the percentage of stocks that are subject to non-synchronous trading problems, namely either suspended or hitting (lower or upper) price limits, within each index. On average, there are 37%, 26% and 10% such index-constituent stocks for the CSI 500, CSI 300 and SSE 50 indices respectively. Overall, figure 1 suggests that there are serious non-synchronous trading issues for the CSI 500 and CSI 300 index-constituent stocks during the market crash. Therefore, the indices may be distorted and so are the futures bases.

To mitigate the distortions of indices caused by stock suspensions or price limits, we recover the 'true' index levels by using the ETFs that track the performances of the corresponding portfolios of underlying indices, as well as the total return indices (TRI) which have the same component stocks as the underlying indices but reflect dividend yields. The recovering logic goes as follows. During our sample period, these ETFs have been continuously traded almost without hitting price limits, hence the log cumulated returns on these ETFs would be good proxies of the 'true' log returns on the TRIs had there been no non-synchronous problems. We then take the differences between the TRIs and the underlying indices to obtain proxies of the dividend yield on each underlying index as both the TRIs and the underlying indices suffer from non-synchronous issues. Lastly, the 'true' log returns of the TRIs from the first step, subtracting the implied dividend yields from the second step, can be viewed as the 'true' cumulated returns on the underlying indices. For brevity, details of how to overcome these issues and to construct the 'true' bases are provided in section 1 of the appendix.

The 'true' basis is defined as $\widehat{Basis}_t^i = \log(F_{t,T}^i) - \log(\hat{S}_t^i)$, where $F_{t,T}^i$ is the futures price of index i with maturity $T-t$ at time t, and $\hat{S}_t^i$ the recovered underlying index level using the above approach. As the cost-of-carry model suggests, no arbitrage implies that $F_{t,T}^i = S_t^i e^{(r_{t,T}-d_{t,T}^i)(T-t)}$ where $r_{t,T}$ and $d_{t,T}^i$ are the risk-free rate and the expected dividend yield of index i from t to T. The basis of index futures i with maturity $(T-t)$ then should simply equal $\log(F_{t,T}^i) - \log(S_t^i) = (r_{t,T} - d_{t,T}^i)(T-t)$.

Note that in general, the dividend payment by Chinese listed firms is concentrated in the period from May to July each year. In addition, according to regulations annual earnings announcements including dividend policy must be disclosed within the first 120 calendar days of an account year. Therefore, the uncertainty on future dividend payments should be low by June when the crash starts. Moreover, the risk-free rate remains positive throughout our sample period. These combined imply that typically we should observe slightly positive bases. However, in our whole sample period the bases have been negative, suggesting that factors other than the interest rate and dividend yield exist.

Therefore, to focus on effects other than the interest rate and dividend yield, in our following analysis, we define backwardation of index futures contract i (denoted by $\widetilde{Bward}_t^i$) as below:

$$\widetilde{Bward}_t^i = \log(F_{t,T}^i) - \log(\hat{S}_t^i) - (r_{t,T} - d_{t,T}^i)(T - t)$$

For comparison purposes and robustness checks, we also calculate the $\widehat{Bward}_t^i$ which does not account for the distortion on backwardation resulted from non-synchronous trading as below:

$$\widehat{Bward}_t^i = \log(F_{t,T}^i) - \log(S_t^i) - (r_{t,T} - d_{t,T}^i)(T - t)$$

3.2. Measuring tail risk aversion

Investors' panic and tail risk aversion are not directly observable but can be inferred through the difference in the premia paid for good uncertainty and bad uncertainty by investors. Drechsler (2013) claims that the variance risk premium (VRP) as first proposed in Carr and Wu (2009) reflects market uncertainty. Miao *et al.* (2019) estimate that 96% of the VRP can be attributed to investors' ambiguity aversion. Orlik and Veldkamp (2014) hold that skewness in macroeconomic variables reflects the uncertainty over 'black swan' events.

Bollerslev and Todorov (2011) propose another indicator of uncertainty, calculated as the difference between the negative jump risk premium and the positive one, and argue that it is a fear measure reflecting the market's premium for extreme uncertainty. Feunou *et al.* (2018) construct the skewness risk premium (SRP) by separating the premia that are attributable to shouldering the upside risk and the downside risk. The SRP is the main indicator of tail risk aversion adopted in our study.† For comparison, we also use the VRP computed from SSE 50 ETF option prices.

Specifically, the upside risk-neutral variance ($IV_t^U = E_t^Q[RV_{t,h}^U(\kappa)]$) is calculated as follows:

$$E_t^Q[RV_{t,h}^U(\kappa)] \approx 2\int_{F_te^{\kappa_F}}^{\infty} \frac{M_0(K)}{K^2}\mathrm{d}K,$$

† We use the SRP to measure the tail risk (uncertainty) premium. This approach is easier to calculate than that in Bollerslev and Todorov (2011). Particularly, the SRP avoids separating the jump and diffusion parts from spot returns under both risk neutral measure and physic measure, which is necessary in Bollerslev and Todorov (2011).

where $M_0(K) = min(P_t(K), C_t(K))$,$P_t(\underline{S})$, $C_t(\underline{S})$, $\underline{S}$ and K are the European put option price, the European call option price, the price and the strike price, respectively. Similarly, we obtain the downside risk-neutral variance $IV_t^D = E_t^Q[RV_{t,h}^D(\kappa)]$ with $E_t^Q[RV_{t,h}^D(\kappa)] \approx 2\int_0^{F_te^{\kappa_F}} \frac{M_0(K)}{K^2}\mathrm{d}K$.

The annualized realized upside variance (RV_t^U) and the annualized realized downside variance (RV_t^Q) of the next 21 trading days are computed as follows:

$$RV_t^U = \frac{251}{21}\Sigma_{i=0}^{21}RV_{t+i}^U, \quad \text{where} \quad RV_t^U = \sum_{j=1}^{n} I_{(r_j\geq 0)}r_{j,t}^2$$

$$RV_t^D = \frac{251}{21}\Sigma_{i=0}^{21}RV_{t+i}^U, \quad \text{where} \quad RV_t^D = \sum_{j=1}^{n} I_{(r_j\leq 0)}r_{j,t}^2$$

where $r_{j,t}$ is the logarithmic difference of the closing prices at trading period j and $j-1$ on trading day t, n is the number of 5-minute trading intervals on trading day t. The skewness of the realized variance is defined as follows:

$$RSV_{t,h}(\kappa) = RV_{t,h}^D(\kappa) - RV_{t,h}^U(\kappa)$$

The SRP is then the difference between the downside variance risk premium and upside variance risk premium:

$$\begin{aligned} SRP_{t,h} &= E_t^Q[RSV_{t,h}] - E_t^P[RSV_{t,h}] \\ &= (IV_t^D - IV_t^U) - (RV_t^D - RV_t^U) \\ &= (IV_t^D - RV_t^D) - (IV_t^U - RV_t^U) \end{aligned}$$

Similar to CBOE, the SSE publishes during our sample period the implied annualized volatility index of SSE 50 ETF options, denoted by iVIX. The VRP on trading day t is computed as:

$$VRP_t = iVIX^2 - RV_t^2,$$

where $RV_t^2 = RV_{t,h}^D(\kappa) + RV_{t,h}^U(\kappa)$ is the annualized realized variance.

3.3. Vector autoregression model specification

To identify the SRP's net effect on normal backwardation and to explain its dynamics, the empirical study is implemented in two stages. In stage 1, an OLS regression is applied to isolate the SRP effect on bases from that of control variables; in stage 2, the residuals of the regression in stage one is used as endogenous variable to implement the VAR analysis.

The first stage model is specified as follows:

$$\widetilde{Bward}_t^i = \alpha_{\widetilde{Bward}}^i + \beta_{\widetilde{Bward}}^i X_t^i + \xi_{\widetilde{Bward},t}^i$$

$$SRP_t = \alpha_{SRP} + \beta_{SRP}X_t^{SRP} + \xi_{SRP,t}$$

where $i \in \{IC, IF, IH\}$, $\alpha_{\widetilde{Bward}}^i$ and α_{SRP} are intercepts, $\beta_{\widetilde{Bward}}^i$ and β_{SRP} are coefficient vectors, X_t^i and X_t^{SRP} are control variable vectors, $\xi_{\widetilde{Bward},t}^i$ and $\xi_{SRP,t}$ are the residuals series.

We specify that $X_t^{SRP} = [r_t^{ETF}, RV_t^{ETF}]'$ where r_t^{ETF} and RV_t^{ETF} are the return and realized volatility of the SSE 50

ETF. Chen *et al.* (1995) show that spot return and volatility affect the futures market hedging positions and the futures basis. There is also potentially a maturity effect that the basis may shrink as the deliver day gets close. Hence, we specify that $X_t^i = [M_t, r_t^i, RV_t^i]'$ where M_t, r_t^i, RV_t^i are futures contract maturity, spot return and spot realized volatility on trading day t.

Roll *et al.* (2007) argue that liquidity changes derived from arbitrage transactions affect basis dynamics and that liquidity and basis are mutually reinforcing. Because there has been a sustained negative basis since China's market crash in 2015, arbitrage transactions could only be achieved by short selling spot and longing futures. However, since August 2015, major brokerage firms have suspended securities margin trading business. Therefore, this type of mutually reinforcing mechanism of liquidity and basis induced by arbitrage transactions did not exist in our sample period and should at most have negligible effect. Despite this, for robustness checks, we consider a control variable set which includes liquidity variables:

$$X_t^i = [M_t, r_t^i, RV_t^i, RS_{s,t}^i, RS_{f,t}^i]'$$

where $i \in \{IC, IF, IH\}$ represents the underlying index, $RS_{s,t}^i$ is the average relative bid-ask spread of index i's component stock, $RS_{f,t}^i$ is the futures contract i's relative bid-ask spread.†

In stage 2, a pair-wise VAR model is adopted with the following settings:

$$Y_t = \alpha + \Sigma_{p=1}^{P} \Phi_p Y_{t-p} + \xi_t$$

where $Y_t = [\xi_{SRP,t}, \xi^i_{\widetilde{Bward},t}]$. Standard Granger causality tests, orthogonal impulse response analysis and forecast error variance decomposition are then implemented. Because the results of orthogonal impulse response analysis and forecast error variance decomposition are well known to be sensitive to the order of endogenous variables, alternative specifications of the order of endogenous variables are used for robustness checks.

3.4. MVMQ-CAViaR model

The VAR model investigates the relationship between the $\xi_{SRP,t}$ and the $\xi^i_{\widetilde{Bward},t}$ at the center. However, those effects can depend on specific market states such as tail event occurrence. To analyze whether the relationship between the futures backwardation and tail risk aversion holds in general or just during the market turmoil, we estimate the tail risk aversion's impact on futures backwardation in different market states (defined as different quantiles) based on the MVMQ-CAViaR model developed by White *et al.* (2015). The MVMQ-CAViaR model can be conveniently thought of as a VAR extension to quantile models. It is flexible to investigate the relationship between economic variables not only at the center but also over the entire conditional distribution of the dependent variable.

Similar to the VAR model specification, we also implement a one-stage OLS regression first to isolate the SRP effect on bases from that of control variables. Then the residuals of the one-stage regression are used as endogenous variables to implement the MVMQ-CAViaR analysis. The specific MVMQ-CAViaR expressions are as follows:

$$\begin{aligned} q_{1t}(\theta) &= c_1(\theta) + a_{11}(\theta)Y_{1t-1} + a_{12}(\theta)Y_{2t-1} \\ &\quad + b_{11}(\theta)q_{1t-1}(\theta) + b_{12}(\theta)q_{2t-1}(\theta) \\ q_{2t}(\theta) &= c_2(\theta) + a_{21}(\theta)Y_{1t-1} + a_{22}(\theta)Y_{2t-1} \\ &\quad + b_{21}(\theta)q_{1t-1}(\theta) + b_{22}(\theta)q_{2t-1}(\theta) \end{aligned}$$

The above formulae can be represented in matrix form as

$$q_t = c + AY_{t-1} + Bq_{t-1}$$

where $q_t(\theta)$ represents the conditional quantile of $[\xi_{SRP,t}, \xi_{\widetilde{Bward}}]'$ under probability θ, which is affected by lagged conditional quantile $q_t(\theta)$ and the original level of vector $Y_t = [\xi_{SRP,t}, \xi_{\widetilde{Bward}}]'$. The elements of matrix A and B are parameters to be estimated. Obviously, the MVMQ-CAViaR model not only captures the shock on quantiles of the endogenous variable from its own lagged level and conditional quantiles but also incorporates the shocks from other time series (or conditional quantiles). Therefore, MVMQ-CAViaR can model the complex dependencies between multiple time series and test whether there is tail dependence among interested variables.

Another advantage of MVMQ-CAViaR is that it allows us to estimate impulse-response functions at different quantiles, from which we can assess the resilience of index futures bases to shocks from tail risk aversion, as well as their persistence. As the standard impulse response analysis where a one-off intervention δ is given, we assume that the one-off intervention δ is given to the error term ϵ_t only at time t. The dynamic behavior of Y_t is described by the VAR model, that is, $Y_t = \sum_{i=1}^{p} \phi_i Y_{t-i} + \epsilon_t$. The time path of ϵ_t is expressed as follows:

The time path of ϵ_{1t} without the intervention would be

$$\ldots, \epsilon_{1t-2}, \epsilon_{1t-1}, \epsilon_{1t}, \epsilon_{1t+1}, \epsilon_{1t+2}, \ldots$$

while the time path with the intervention would be

$$\ldots, \epsilon_{1t-2}, \epsilon_{1t-1}, \tilde{\epsilon}_{1t}, \epsilon_{1t+1}, \epsilon_{1t+2}, \ldots$$

The θ th-quantile impulse-response function (QIRF) for the ith error term denoted as $\Delta_{i,s}(\tilde{\epsilon}_{1t})$ is defined as

$$\Delta_{i,s}(\tilde{\epsilon}_{1t}) = \tilde{q}_{i,t+s} - q_{i,t+s}, \quad s = 1, 2, 3, \ldots$$

† The relative spread of spot, RS_s^i, is calculated by $RS_{s,t}^i = \frac{1}{N}\sum_{j=1}^{N}(\frac{1}{K}\sum_{k=1}^{K}\frac{Bid_k^j - Ask_k^j}{P_t^j})$, where $i \in \{CSI500, CSI300, SSE50\}$, Bid_k^j and Ask_k^j are bid and ask prices of index i's component stock j at intraday time k of date t, respectively. P_k^j is the middle point of bid and ask price of component stock j at time k. N is the number of component stocks in index i and K is the number of samples for price within each intra-day. The relative spread of futures, $RS_{f,t}^i$, is calculated by $RS_{f,t}^i = \frac{1}{K}\sum_{k=1}^{K}\frac{Bid_k^i - Ask_k^i}{P_k^i}$, Bid_t^i and Ask_t^i are bid and ask prices of futures contract i at time k, respectively. P_k^i is the middle point of bid and ask price of futures contract i at time k.

where $\tilde{q}_{i,t+s}$ is the θ th-conditional quantile of the affected series and $q_{i,t+s}$ is the θ th-conditional quantile of the unaffected series.†

4. Empirical findings

4.1. Data and summary statistics

The stocks and futures data are from CSMAR, a major financial data provider in China. We obtain 1-month repo rates in the Chinese interbank market from Wind, another data service provider in China, and use them as risk-free rates. The sample period is from 26th June to 30th September 2015.

Because the A-share market closes at 15:00 and the index futures market closes at 15:15, we use the price level of index, ETF, and futures at 15:00 to calculate the bases. In consideration of the fact that nearest maturity contracts are typically most actively traded and there are many roll-over trades in the last week before delivery, we use nearest maturity futures contracts with maturity between 7 and 60 days to rollover the contracts and obtain the daily time series of futures prices.

Similarly, we select nearest maturity SSE 50 ETF option contracts with a term between 7 and 60 days, take out those extreme outliers that violates the classic option price theoretical boundaries, and calculate the implied volatilities using the Black–Scholes–Merton formula. These implied volatilities are then interpolated with cubic splines with the ETF spot price as the center. The risk-neutral upside and downside variances are calculated using the discrete approximation method.

Table 1 reports the summary statistics of the *Basis*, $\widehat{Basis}$, $\widehat{Bward}$, $\widetilde{Bward}$ and the control variables used in the empirical study. Throughout the sample period, the average bases of three index futures are all negative. The bases of nearest maturity IC, IF, and IH contracts with maturity larger than 7 calendar days average -5.24%, -3.90% and -2.71%, respectively. For $\widehat{Basis}$, the average values are -5.06%, -3.56% and -2.30% respectively, implying that the average annual hedging costs for a typical trader would be close to 60.72%, 42.72% and 27.6%! The difference between *Basis* and $\widehat{Basis}$ are minimal, suggesting that non-synchronous trading can only explain a small fraction of the negativity of the bases. The average $\widetilde{Bward}$ are -5.21%, -3.61% and -2.30% respectively, slightly larger than $\widehat{Basis}$ in absolute value, suggesting that on average, the risk-free rate is a little larger than the expected dividend yield in the sample period. In other words, the high absolute values of average $\widetilde{Bward}$ implies that the difference of risk-free rate and dividend yield is not enough to explain the phenomenal negative bases. It is clear from Panel 1c and Panel 1d that for each contract, $\widehat{Bward}_t^i$ is always larger than $\widetilde{Bward}_t^i$, but the differences are negligible, implying that the non-synchronous trading may have weak explanation power toward backwardation.

Figure 2 demonstrates the dynamics of $\widehat{Basis}_t^i$, $\widehat{Bward}_t^i$ and $\widetilde{Bward}_t^i$ and their behavior with respect to the original basis series. The upper two subfigures show that the time variations of $\widehat{Basis}$ and $\widetilde{Bward}$ are remarkably remain negative in the sample period. The middle right, lower left and lower right subfigures highlight their differences from the original basis series. If non-synchronous trading is not a reason for the negative bases, we should observe little variation in the difference between $\widehat{Basis}$ and the original basis from middle right subfigure. Similarly, because of the very minimal variation of the absolute differences between the risk-free rate and dividend yield during our sample period, the cost-of-carry hypothesis would predict that the $\widetilde{Bward}$ should not be significantly different from the original series either.

It is clear from figure 2 that the non-synchronous hypothesis, but not the cost-of-carry hypothesis, does have some, although not much, explanatory power on the variation of the bases from 26th June to 30th September 2015. If we extend the sample period to 31st December, the explanatory power of the non-synchronous hypotheses is evidently reduced for all three contracts, although it seems to have stronger effect for the IC contracts than for the other two. This is consistent with figure 1, where the IC contracts are associated with more suspensions and price limits between October and December. With these observations, $\widetilde{Bward}$ is used in the following empirical analysis so we can focus on those parts of the bases unexplained by both nonsynchronous trading and cost-of-carry hypotheses.‡

All three index futures are in deep backwardation in our sample period. The IC's backwardation is the deepest among three futures, followed by IF, then IH. The standard deviation (S.D.), maximum, minimum values and volatilities of these three bases series are similarly ranked. Column 7 of table 1 reports the correlations of these variables with SRP. For all indices, bases are highly associated with SRP, with correlation coefficients of -0.63, -0.61 and -0.55 respectively, suggesting that the SRP could be potentially accountable for the backwardation of futures. Due to the endogeneity concern, the causality, instead of the contemporaneous, relationships between SRP and bases will be discussed in a VAR model setting.

Panel 2 in table 1 reports the summary statistics of SRP and iVIX. The average of SRP is 15.20%, indicating that the downside variance risk premium is higher than the upside one. The maximum value of SRP is 112.34%, suggesting that investors is deeply concerned about market tail risk and are willing to pay very high option premium to hedge away the crash risk. The correlation of iVIX with SRP is 0.63, implying that there could be difference between iVIX and SRP of explanatory power toward backwardation.

Figure 3 presents the construction of SRP and the differences between SRP and other uncertainty measurements. The upper left subfigure shows that there is no significant difference between realized downside and upside variances. The upper right subfigure, however, shows that the risk-neutral

† Technical details of the derivation of IRF can be found in White *et al.* (2015) for reference and codes in this paper are available on the website of the journal.

‡ Our results are robust to substituting the $\widehat{Basis}$ and the original basis for $\widetilde{Bward}$.

Table 1. Summary statistics.

	Size	Mean	S.D.	Max	Min	Correlation with SRP
Panel 1: Basis						
Panel 1a: $Basis_t^i$ (%)						
$Basis_t^{IC}$	66	− 5.24	3.39	1.52	− 15.07	− 0.63
$Basis_t^{IF}$	66	− 3.90	2.62	0.60	− 12.72	− 0.61
$Basis_t^{IH}$	66	− 2.71	2.02	0.04	− 10.49	− 0.55
Panel 1b: $\widehat{Basis}_t^i$ (%)						
$\widehat{Basis}_t^{IC}$	66	− 5.06	2.75	0.64	− 13.05	− 0.60
$\widehat{Basis}_t^{IF}$	66	− 3.56	2.57	0.70	− 11.87	− 0.56
$\widehat{Basis}_t^{IH}$	66	− 2.30	1.96	− 0.10	− 9.70	− 0.55
Panel 1c: $\widehat{Bward}_t^i$ (%)						
$\widehat{Bward}_t^{IC}$	66	− 5.39	3.41	1.26	− 15.21	− 0.63
$\widehat{Bward}_t^{IF}$	66	− 3.95	2.67	0.80	− 12.82	− 0.60
$\widehat{Bward}_t^{IH}$	66	− 2.71	2.12	0.47	− 10.56	− 0.52
Panel 1d: $\widetilde{Bward}_t^i$(%)						
$\widetilde{Bward}_t^{IC}$	66	− 5.21	2.77	0.59	− 13.20	− 0.60
$\widetilde{Bward}_t^{IF}$	66	− 3.61	2.63	0.79	− 11.97	− 0.54
$\widetilde{Bward}_t^{IH}$	66	− 2.30	2.05	0.55	− 9.77	− 0.52
Panel 2: SRP and iVIX (Annualized) (%)						
SRP	66	15.19	19.31	112.34	− 1.75	1.00
iVIX	66	44.64	9.41	63.78	30.41	0.63
Panel 3: Control Variables (Daily)						
Panel 3a: Maturity						
$Maturity_t$	66	20.71	8.84	39.00	7.00	− 0.04
Panel 3b: Underlying Return (%)						
r_t^{50ETF}	66	− 0.38	3.55	8.09	− 10.52	− 0.09
r_t^{IC}	66	− 0.61	4.16	6.39	− 8.73	− 0.31
r_t^{IF}	66	− 0.49	3.51	6.50	− 9.15	− 0.19
r_t^{IH}	66	− 0.39	3.47	7.55	− 9.85	− 0.12
Panel 3c: Realized Volatility (%)						
RV_t^{50ETF}	66	3.19	2.22	10.87	0.96	0.41
RV_t^{IC}	66	3.71	1.74	8.82	1.42	0.36
RV_t^{IF}	66	3.29	1.88	9.96	1.08	0.38
RV_t^{IH}	66	3.15	2.02	9.53	0.95	0.43
Panel 3d: Average Relative Bid-Ask Spread of Component Stocks (%)						
$RS_{s,t}^{IC}$	66	0.15	0.03	0.24	0.10	0.48
$RS_{s,t}^{IF}$	66	0.13	0.02	0.17	0.09	0.51
$RS_{s,t}^{IH}$	66	0.12	0.02	0.17	0.09	0.64
Panel 3e: Relative Bid-Ask Spread of Futures (%)						
$RS_{f,t}^{IC}$	66	0.07	0.05	0.31	0.02	0.13
$RS_{f,t}^{IF}$	66	0.03	0.03	0.20	0.01	0.15
$RS_{f,t}^{IH}$	66	0.05	0.05	0.27	0.02	0.11

This table reports summary statistics of the sample from 26th June 2015 to 30th September 2015. We use daily data to calculate the returns, intraday 5-minute returns to compute realized volatility, slices of quotations of stocks and futures contracts to obtain relative bid-ask spreads. The frequencies of quotation slices are half second, 3 seconds and 5 seconds for futures contracts in CFFEX, stocks in Shenzhen Stock Exchange, and stocks in Shanghai Stock Exchange, respectively.

downside variance is significantly larger than upside one. This results in the significant difference between downside and upside semi-variance risk premium as shown in the lower left subfigure. Thus the SRP is mainly driven by the difference between risk neutral semi-variances, not that between the realized ones.

The lower right subfigure compares the time series of three uncertainty measures: SRP, $iVIX^2$ and VRP. The SRP is more volatile and has more extreme values than the VRP and $iVIX^2$. During the market crash, both SRP and $iVIX^2$ rose sharply, whereas the VRP declined abruptly, which is consistent with Bollerslev *et al.* (2011) who note that during the period of severe market volatility, negative variance risk premium may emerge. In addition, the dynamic of SRP fits the market crash development better than both VRP and $iVIX^2$: it is quite volatile during the crash, with peaks corresponding to the most dramatic price drops, and then quickly returns to normal afterwards. Graphically speaking, the SRP seems to be more consistent with the dynamics of bases than either $iVIX^2$ or VRP.

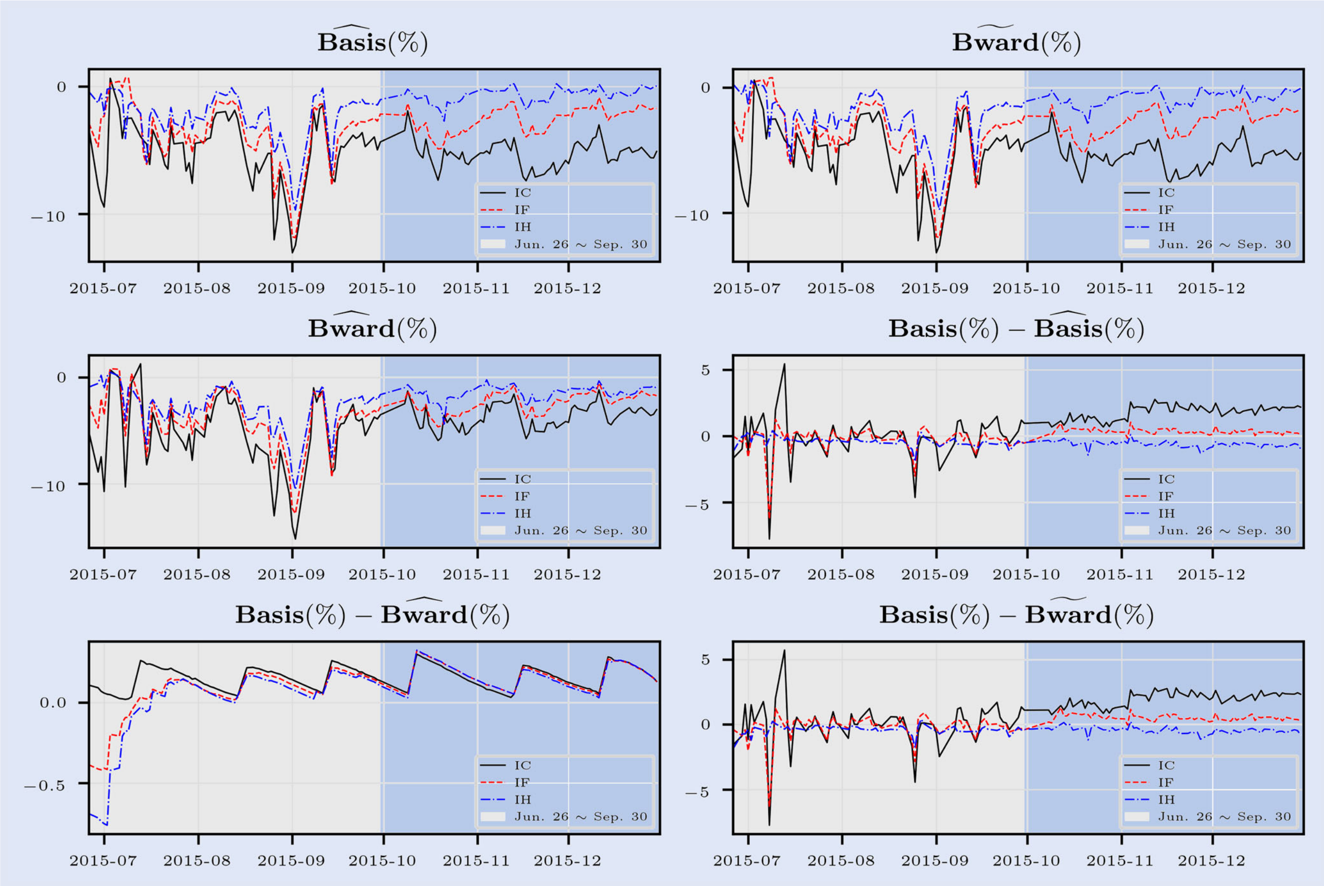

Figure 2. $\widehat{Basis}_t^i$, $\widehat{Bward}_t^i$, $\widetilde{Bward}_t^i$.

Time series of $\widehat{Basis}_t^i$, $\widehat{Bward}_t^i$, $\widetilde{Bward}_t^i$ are presented. The greyed regions correspond to the period from 26th June 2015 to 30th September 2015.

Panel 3 of table 1 reports the summary statistics of control variables. The maturities of futures contracts, as shown in Panel 3a, are from 7 to 39 calendar days, with an average of 20.71. Panel 3b reports the SSE 50 ETF returns and index returns. In our sample period, all returns have negative averages, with the CSI 500 index having the lowest average daily return of -0.61%. Among three indices, the return on the CSI 500 index has the highest absolute correlation of 0.31 with SRP. Panel 3c reports the summary statistics of monthly realized volatility calculated with 5-minute returns. The CSI 500 index has the highest realized volatility of 3.71%, followed by the CSI 300, then the SSE 50 index. But the volatility of volatility and the maximum of realized volatility show the opponent rankings in mean. The realized volatilities all show modest correlations (around 0.40) with SRP. Panel 3d reports the average relative bid-ask spread on index-constituent stocks. The CSI 500 constituent stocks have the highest mean (0.15%) of average relative bid-ask spread among three indices. The standard deviation, maximum and minimum values exhibit similar rankings, together indicating that the CSI 500 market has the worst liquidity among the three markets under examination. The average relative bid-ask spread of futures has lower mean, higher standard deviation, lower correlations with SRP than index-constituent stocks.

4.2. *Can return and volatility explain the sustained backwardation?*

Chen *et al.* (1995) argue that return and volatility are major factors in explaining the futures basis in the U.S. market. To test this hypothesis in the 2015 Chinese stock market crash, following Chen *et al.* (1995) and Roll *et al.* (2007), we first regress the futures backwardations on the maturity of futures contracts, daily index return, realized volatility and the relative bid-ask spread of the underlying. To isolate the influence of ETF volatility and return on SRP, the SRP is also regressed on these two variables.

The regression results are reported in table 2. The coefficients of maturity are negative and significant in all regressions, whether the relative bid-ask spread variables are included, indicating that bases always converge to zero as futures contracts get close to its expiration day. More importantly, in sharp contrast with Chen *et al.* (1995), whether the underlying or the futures relative bid-ask spread is controlled or not, there is no significant impact of volatility and return on the bases. In addition, a comparison of the adjusted R^2s between models including relative bid-ask spreads with those excluding the spreads reveals that the liquidity variables have just marginal explanatory power for our sample.

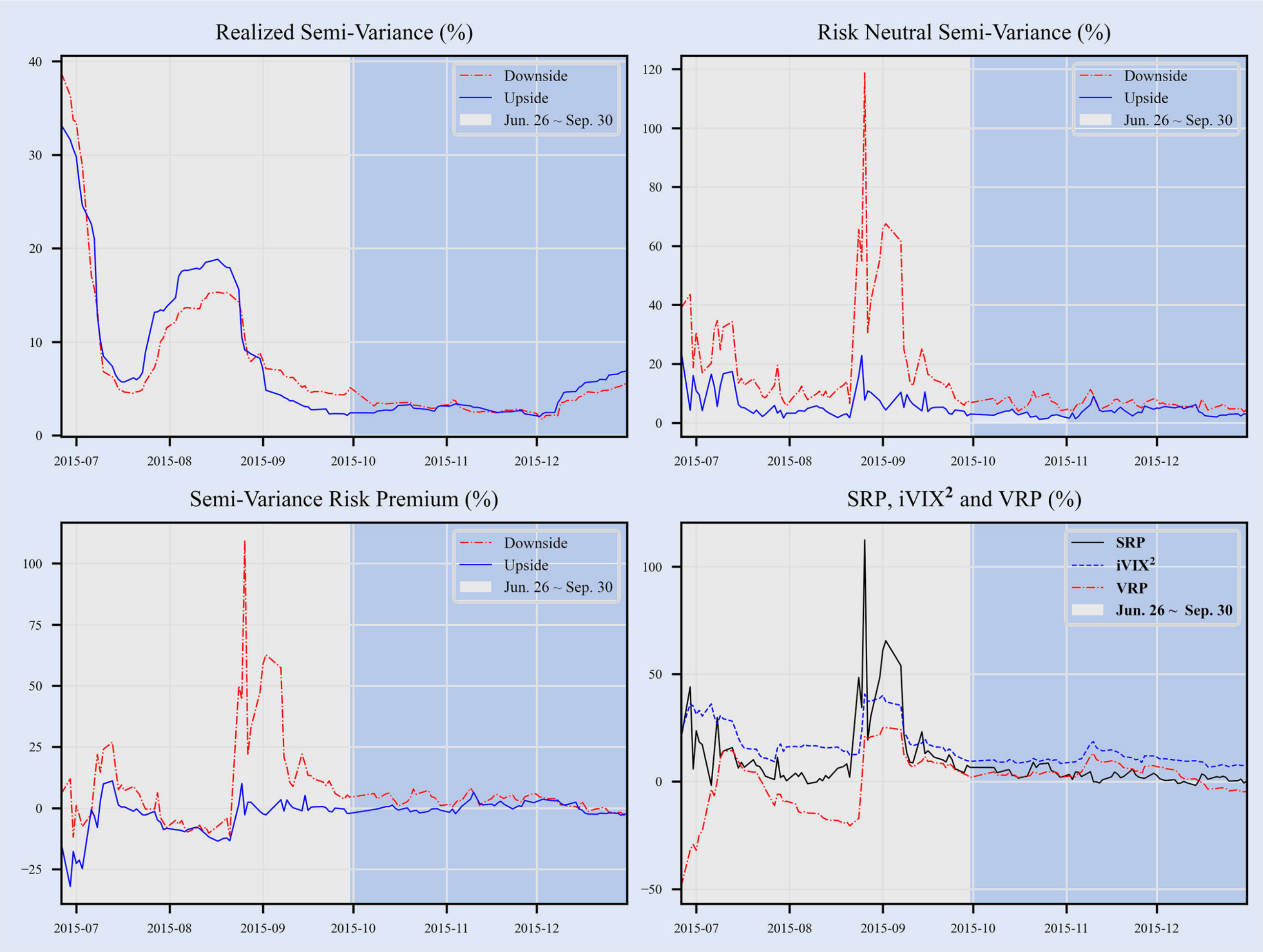

Figure 3. Construction of skewness of variance risk premium.
The greyed regions correspond to the period from 26th June 2015 to 30th September 2015.

These indicate that the model proposed by Chen *et al.* (1995) and the arbitrage trading mechanism in Roll *et al.* (2007) are neither sufficient to explain the deep backwardation in the Chinese index futures during the market crash in 2015. A new model or a new set of variables are warranted to explain this phenomenon.

The last column of table 2 reports the regression results on SRP. The coefficient of the ETF volatility is positive and significant at the 10% level, suggesting that the SRP is increasing with ETF realized volatility.

4.3. The VAR analysis

During a stock market collapse, hedging demand can be related to market volatility as pointed out in the literature. However, it could also be related to hedgers' concern over uncertainty in the future, which can be proxied by SRP. Many stocks that hit the daily lower price limit will trigger a market liquidity shortage, which in turn will further prompt investors to lower their expectations about the future stock price movement. To make it even worse, due to restrictions on short selling and spot sale, investors may rush to the door by selling index futures, only to depress the futures prices further. Therefore, the futures basis should be negatively associated with hedging demand, which is proxied by the market skewness risk premium. In this section, we test the lead-lag relationship between SRP and the futures bases using a VAR model.

4.3.1. Granger-causality between SRP and $\widetilde{Bward}$. When the market is in panic, hedging demand using index futures increases, hence we would expect that a rising SRP may lead to a deepening backwardation. The other way around may also be true. Option market makers and other investors often use index futures to hedge their option positions, hence the implied downward variance premium of options may rise along with the deepening backwardation of index futures.

To address the lead-lag relationship between endogenous variables, based on the VAR setting, a Granger causality test is conducted on each of the endogenous variables. Table 3 reports the Wald statistics of the pair-wise Granger causality test. At the 10% (1%) significance level, the SRP Granger-causes the $\widetilde{Bward}$ of the present-month IC(IF) contracts; but it is not the Granger cause of the basis of present-month IH contracts. On the other hand, except for IH which is less in backwardation, $\widetilde{Bward}$ does not Granger-cause the SRP.

Nevertheless, one must be careful to interpret the test results of Granger causality between endogenous variables. In a pair-wise Granger causality test, the correlation between tested variables may be reflected in the disturbance term of

Table 2. Stage 1 regression results.

	$\widetilde{Bward}_t^{IC}$	$\widetilde{Bward}_t^{IF}$	$\widetilde{Bward}_t^{IH}$	$\widetilde{Bward}_t^{IC}$	$\widetilde{Bward}_t^{IF}$	$\widetilde{Bward}_t^{IH}$	SRP
Intercept	− 1.52	− 0.59	0.44	0.58	2.26	3.42	3.61
	(0.97)	(0.73)	(0.45)	(2.96)	(2.81)	(2.05)	(4.07)
$Maturity_t$	− 0.10**	− 0.14***	− 0.10***	− 0.13***	− 0.15***	− 0.11***	
	(0.04)	(0.03)	(0.02)	(0.04)	(0.03)	(0.01)	
r_t^i	0.08	− 0.02	− 0.04	0.06	− 0.01	− 0.03	
	(0.08)	(0.08)	(0.07)	(0.07)	(0.08)	(0.07)	
RV_t^i	− 0.42	− 0.03	− 0.20	− 0.18	0.20	− 0.04	
	(0.28)	(0.25)	(0.18)	(0.31)	(0.27)	(0.16)	
$RS_{s,t}^i$				− 19.68	− 27.53	− 28.51	
				(26.68)	(30.39)	(20.95)	
$RS_{f,t}^i$				7.49	− 1.67	− 0.31	
				(7.72)	(11.24)	(5.11)	
r_t^{50ETF}							− 0.44
							(0.65)
RV_t^{50ETF}							3.57*
							(1.75)
Adj. R^2	0.18	0.18	0.16	0.18	0.19	0.16	0.15

This table reports estimators of the stage 1 regression using the sample from 26th June 2015 to 30th September 2015. The regression is specified as follows: $\widetilde{Bward}_t^i = \alpha^i_{\widetilde{Bward}} + \beta^i_{\widetilde{Bward}} X_t^i + \xi^i_{\widetilde{Bward},t}$, $i \in \{IC, IF, IH\}$. For the second, third and fourth columns, $X_t^i = [M_t, r_t^i, RV_t^i]'$; for the fifth, sixth and seventh columns, $X_t^i = [M_t, r_t^i, RV_t^i, RS_{s,t}^i, RS_{f,t}^i]'$; for the last column, $X_t^{SRP} = [r_t^{ETF}, RV_t^{ETF}]'$. Standard deviations are reported in parentheses which are based on Newey–West corrections. *, **, and *** denote statistical significance at 10%, 5% and 1% levels, respectively.

Table 3. Granger causality tests for bivariate VAR with $\xi_{SRP,t}$ and $\xi^i_{\widetilde{Bward},t}$.

	$\widetilde{Bward}_t^{IC}$	$\widetilde{Bward}_t^{IF}$	$\widetilde{Bward}_t^{IH}$
H_0: the row variable does not Granger-cause column variables			
SRP	5.69*	16.99***	2.45
H_0: the column variable does not Granger-cause row variables			
SRP	2.54	8.88	4.19

This table reports the Wald statistics of Granger Causality tests of bivariate VAR with $\xi_{SRP,t}$ and $\xi^i_{\widetilde{Bward},t}$ using the sample from 26th June 2015 to 30th September 2015. *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively.

the model, and the contemporaneous correlation in the disturbance term enters the next phase through the lag term of the regression, exerting influences on other endogenous variables. Therefore, significance in the Granger causality test suggests that the explanatory variables must contain information on the predictable explained variables, whereas insignificance in the Granger causality test does not necessarily mean that the explanatory variables do not contain information about the predictable explained variables. An orthogonalized impulse response analysis is performed next to address these issues.

4.3.2. Impulse response analysis. Impulse response analysis is commonly applied to analyze the impact attributable to a certain variable on other variables in a later phase through the Cholesky decomposition of the variance-covariance matrix of the disturbance terms. A well-known problem with the variance decomposition of impulse response analysis and prediction is its variable-order-sensitivity. Results presented in this section are based on the variable order of (SRP, Basis), as the options market may have played a better role of price discovery than the futures market during the market turmoil in 2015 for the following reasons: first, according to the SSE, the trading of institutional investors accounted for 79.2% of the total turnover of the SSE 50 ETF option market in 2015, and institutional investors on average have information advantage than retail investors popular in the futures markets; second, the index futures trading restrictions imposed by the CFFEX forced a large number of investors to leave the market, which supposedly hindered the functioning of the index futures market. In contrast, the trading volume of SSE 50 ETF options had been steadily increasing during the same period, implying that the option prices might incorporate more information than the futures prices; and third, market makers in the option market might have played a positive role in price discovery, in particular during the market crash.

Figures 4– 6 present respectively the results of the orthogonal impulse response analysis for the pair $[\xi_{SRP,t}, \xi^i_{\widetilde{Bward},t}]$ where $i \in \{IC, IF, IH\}$. Because $\xi^i_{\widetilde{Bward},t}$ is part of $\widetilde{Bward}_t^i$, in the following discussions $\widetilde{Bward}_t^i$ is used to substitute for $\xi^i_{\widetilde{Bward},t}$.

The upper left and lower right subfigures of figures 4–6 show that shocks on both SRP and $\widetilde{Bward}_t^i$ are persistent. The upper right subfigure of figure 5 shows that the $\widetilde{Bward}_t^{IF}$ are informative in predicting future SRP. One standard deviation increase of $\widetilde{Bward}_t^{IF}$ will result in a 4.55% decrease of SRP in the following period. The decrease of SRP is significant at the 5% level. For $\widetilde{Bward}_t^{IC}$ and $\widetilde{Bward}_t^{IH}$, the responses are negative but insignificant.

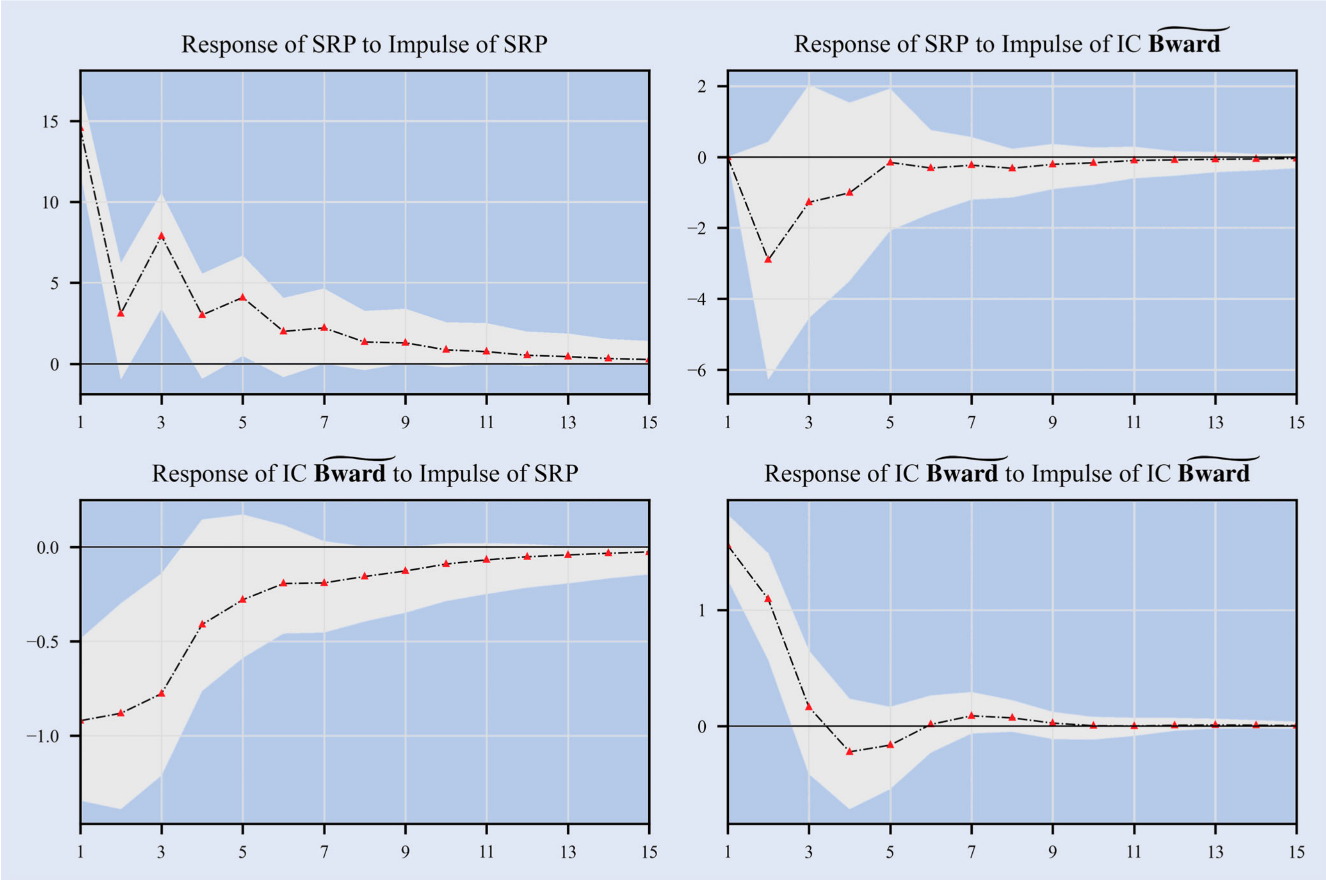

Figure 4. Impulse response functions for the bivariate VAR with SRP_t and $\widetilde{Bward}_t^{IC}$.
Impulse response functions of bivariate VAR with $\xi_{SRP,t}$ and $\xi^{IC}_{\widetilde{Bward},t}$ estimated in sample from 26th June 2015 to 30th September 2015 are presented. The $\xi^{IC}_{\widetilde{Bward},t}$ and $\xi_{SRP,t}$ are obtained in the first-stage regression with specifications of $X_t^{SRP} = [r_t^{50\ ETF}, RV_t^{50\ ETF}]'$ and $X_t^{IC} = [M_t, r_t^{IC}, RV_t^{IC}]'$, respectively. The upper left subfigure is the response of the $\xi_{SRP,t}$ to one standard deviation shock of $\xi_{SRP,t}$. The upper right subfigure is the response of the $\xi_{SRP,t}$ to one standard deviation shock of $\xi^{IC}_{\widetilde{Bward},t}$. The lower left subfigure is the response of $\xi^{IC}_{\widetilde{Bward},t}$ to one standard deviation shock of $\xi_{SRP,t}$, and the lower right subfigure is the response of $\xi^{IC}_{\widetilde{Bward},t}$ to one standard deviation shock of $\xi^{IC}_{\widetilde{Bward},t}$. The greyed area is the 95% confidence interval calculated by Monte Carlo simulation (based on 10,000 simulations).

SRP is informative in predicting future $\widetilde{Bward}_t^i$. For all futures contracts, the lower left subfigures show that shocks on SRP have negative and significant impact on $\widetilde{Bward}_t^i$ in the following two or three periods. A standard deviation shock on SRP will result in a 0.96%, 0.74% and 0.64% decrease of $\widetilde{Bward}$ contemporaneously for IC, IF and IH respectively. Those effects are significant in the following two periods for IC, and one period for IF and IH.

4.3.3. Variance decomposition of endogenous variables.

Table 4 reports the results of the variance decomposition of the basis and the SRP of each present-month contract as another way of describing the dynamics of basis and SRP. We divide table 4 into three panels to report the results of forecasted errors variance decomposition for endogenous variables $[\xi_{SRP,t}, \xi^{IC}_{\widetilde{Bward},t}]$,$[\xi_{SRP,t}, \xi^{IF}_{\widetilde{Bward},t}]$ and $[\xi_{SRP,t}, \xi^{IH}_{\widetilde{Bward},t}]$, respectively. The second row lists the forecasted variable, the third row lists variables that are used to predict, and the rest rows report the ratios of variance that can be attributed to the variables in the third row. Overall, results of longer-than-15 phases exhibit convergence characteristics.

Columns 1 and 2 of Panel 4a show that more than 97% variance on predicted errors of SRP from period 1 to 15 can be attributed to its own shock, while the rest 3% can be attributed to $\widetilde{Bward}_t^{IC}$. In comparison, from column 3 and column 4, the proportion of variance on predicted errors of $\widetilde{Bward}_t^{IC}$ from periods 1 to 15 that can be attributed to SRP ranges from 29% to 44%, and that attributable to its own shock ranges from 71% to 56%.

Results in Panel 4b and Panel 4c are similar to the above. The major differences are: for $\xi^{IF}_{\widetilde{Bward},t}$, there are even more variations which can be attributed to SRP than $\xi^{IC}_{\widetilde{Bward},t}$, and for $\xi^{IH}_{\widetilde{Bward},t}$ the variance of predicted errors that can be attributed to SRP ranges between 21% and 35%, which is less than both $\xi^{IC}_{\widetilde{Bward},t}$ and $\xi^{IF}_{\widetilde{Bward},t}$.

Figure 5. Impulse response functions for the bivariate VAR with SRP_t and $\widetilde{Bward}_t^{IF}$.
The model specification to obtain this figure is similar with figure 4.

The above demonstrates that shocks on SRP can explain a significant proportion of the following period variations of futures backwardations, but not vice versa. All the way to phase 15, 21% –57% of the forecast error variance can be attributed to the disturbance of the SRP. Across the contracts, the explanatory power of SRP on $\widetilde{Bward}_t^{IC}$ and $\widetilde{Bward}_t^{IF}$ is stronger than that on $\widetilde{Bward}_t^{IH}$, which is consistent with the Granger causality test and the impulse response analysis based on the VAR model. These imply that the futures backwardation does partially reflect the risk premium paid by futures traders to hedge the tail risk, and it also represents the 'insurance premium' requested by the similarly tail risk-averse speculators in providing risk management services for hedgers.

4.4. *MVMQ-CAViaR analysis*

To differentiate the impacts of tail risk aversion on the index futures backwardation across market conditions, we extend our sample from 26th June 2015 to 29th December 2017 (the last trading date of 2017) which includes 571 observations covering normal and volatile market periods to estimate the MVMQ-CAViaR model. Since both SSE 50 and CSI 300 ETFs have paid cash dividends on 28th November 2016 and 19th January 2016, and from figure 1, the number of firms that suffered non-synchronous trading problems has greatly decreased, we decide to use the original futures backwardation for IH and IF contracts from ex-dividend date to the last trading day of the extended sample without adjusting for the distortions as we do in previous sections.

The one-stage regression results for the extended sample are reported in table 5. The coefficients of maturity are negative and significant in all regressions. In contrast with the regression results during the market crash in 2015, there is significantly negative impact of volatility on the bases, suggesting that the bases are decreasing with index realized volatility, consistent with Chen *et al.* (1995). It is worth noting that, although relative bid-ask spreads have marginal explanatory power during the market crash, the underlying and futures relative bid-ask spreads have significant impact on bases in the extended sample. The incremental improvement in the R^2 associated with adding the relative bid-ask spreads indicates that the liquidity effect can substantially help explain the sustained backwardation. Therefore, when we use the residuals derived from one-stage regressions and estimate the MVMQ-CAViaR model, we explicitly control these variables.

Note that high SRP implies high tail risk premium, and high absolute negative basis corresponds to deep backwardation, however, the MVMQ-CAViaR model requires to specify the same quantile for each endogenous variable. Thus we take $[-\xi_{SRP,t}, \xi^i_{\widetilde{Bward},t}]'$ to estimate MVMQ-CAViaR model and report the negative impulse response functions of $-\xi_{SRP,t}$ in order to compare directly with those results of the VAR model. And the 0.1-quantile, 0.5-quantile and 0.9-quantile

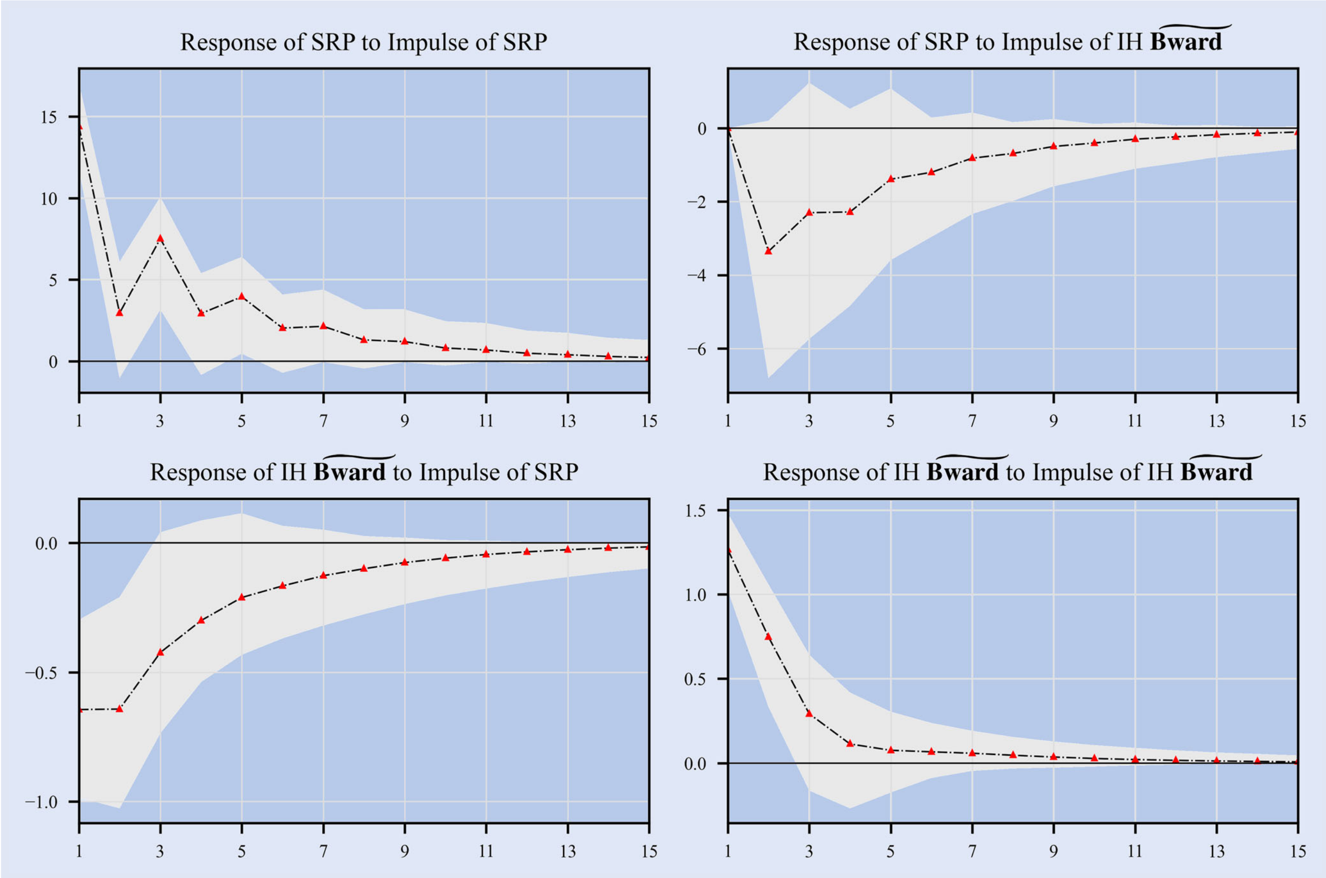

Figure 6. Impulse response functions for the bivariate VAR with SRP_t and $\widetilde{Bward}_t^{IH}$. The model specification to obtain this figure is similar with figure 4.

correspond to the volatile, normal and tranquil markets, respectively.

Figures 7– 9 illustrate the estimated response surfaces with respect to quantiles and persistent periods from the MVMQ-CAViaR model for the endogenous variables pair $[\xi_{SRP,t},\ \xi^{i}_{\widetilde{Bward},t}]'$, where $i \in \{IC, IF, IH\}$, respectively.

The left subfigure of figure 7 presents the response function of $\widetilde{Bward}_t^{IC}$ to a shock from SRP_t. It is evident that the response of $\widetilde{Bward}_t^{IC}$ to the shock from SRP_t depends on the quantile. At lower quantiles, such as the 0.1 quantile, the impulse of SRP_t generates larger negative responses of $\widetilde{Bward}$ than at higher quantiles, suggesting that high tail risk premium can result in deep backwardation of index futures. At higher quantiles, such as the 0.9-quantile, the response of $\widetilde{Bward}_t^{IC}$ to a shock of SRP_t becomes positive.

The right subfigure of figure 7 shows the response function of SRP_t to a shock from $\widetilde{Bward}_t^{IC}$. The response of SRP_t is relatively flat across quantiles. Compared to the left subfigure, at the same quantile, there are larger negative responses of SRP_t to one standard deviation positive shocks from $\widetilde{Bward}_t^{IC}$ within about 10 periods, which implies that a decreasing backwardation may lead to lower tail risk premium. Figures 8 and 9 exhibit similar patterns.

Next, we turn to analyze the significance of the impulse response functions. Figures 10–12 present the impulse-response functions and their 95% confidence intervals for the pair $[\xi^{i}_{\widetilde{Bward},t}, \xi_{SRP,t}]$ where $i \in \{IC,\ IF,\ IH\}$. The upper left subfigures of figures 10–12 indicate that for the IC, IF and IH contracts, under $\theta = 0.1$, one standard deviation increase of SRP will result in a 0.21%, 0.10% and 0.08% decrease of $\widetilde{Bward}$ in the following one or three periods, and the impulse responses are significant in statistics, except for the IH contract. The upper center subfigures reveal that there are weak responses of $\widetilde{Bward}$ to one standard deviation shock from SRP during the normal market. The lower subfigures of figures 10–12 show that $\widetilde{Bward}$ are informative in predicting future SRP for all futures contracts at different quantiles.†

As the impulse response analysis of the MVMQ-CAViaR is order-sensitive, alternative endogenous variable order is used to test the robustness of the analysis.‡ The significance of the response of $\widetilde{Bward}_t^{i}$, $i \in \{IC, IF\}$ to shocks from SRP_t in the left tail remains.

In summary, the empirical findings from the MVMQ-CAViaR model are consistent with those findings in the VAR analysis. The SRP can explain a significant proportion of the variations of futures backwardation in the following periods.

† These results may seem in contrast with those from the VAR analysis in which this relationship is insignificant. It may be related to that the explainability of liquidity to backwardation depends on market conditions which has not been considered in the one stage regression.
‡ These results are reported in the Appendix.

Table 4. Variance decomposition of forecast errors for bivariate VAR with $\xi_{SRP,t}$ and $\xi^{i}_{\widetilde{Bward},t}$.

	Panel 4a				Panel 4b				Panel 4c			
	$\xi_{SRP,t}$		$\xi^{IC}_{\widetilde{Bward},t}$		$\xi_{SRP,t}$		$\xi^{IF}_{\widetilde{Bward},t}$		$\xi_{SRP,t}$		$\xi^{IH}_{\widetilde{Bward},t}$	
	$\xi_{SRP,t}$	$\xi^{IC}_{\widetilde{Bward},t}$	$\xi_{SRP,t}$	$\xi^{IC}_{\widetilde{Bward},t}$	$\xi_{SRP,t}$	$\xi^{IF}_{\widetilde{Bward},t}$	$\xi_{SRP,t}$	$\xi^{IF}_{\widetilde{Bward},t}$	$\xi_{SRP,t}$	$\xi^{IH}_{\widetilde{Bward},t}$	$\xi_{SRP,t}$	$\xi^{IH}_{\widetilde{Bward},t}$
1	1.00	0.00	0.29	0.71	1.00	0.00	0.26	0.74	1.00	0.00	0.21	0.80
2	0.97	0.03	0.34	0.66	0.91	0.09	0.29	0.71	0.95	0.05	0.28	0.72
3	0.97	0.03	0.41	0.59	0.90	0.10	0.31	0.69	0.94	0.06	0.31	0.69
4	0.97	0.03	0.42	0.58	0.90	0.10	0.32	0.68	0.93	0.07	0.34	0.67
5	0.97	0.03	0.43	0.57	0.90	0.10	0.37	0.63	0.93	0.07	0.34	0.66
6	0.98	0.02	0.43	0.57	0.91	0.09	0.49	0.51	0.92	0.08	0.34	0.66
7	0.98	0.02	0.43	0.57	0.91	0.09	0.52	0.48	0.92	0.08	0.34	0.66
8	0.98	0.02	0.44	0.56	0.91	0.09	0.53	0.47	0.92	0.08	0.35	0.66
9	0.98	0.02	0.44	0.56	0.91	0.09	0.55	0.45	0.92	0.08	0.35	0.65
10	0.98	0.02	0.44	0.56	0.91	0.09	0.56	0.44	0.92	0.08	0.35	0.65
11	0.98	0.02	0.44	0.56	0.91	0.09	0.56	0.44	0.92	0.08	0.35	0.65
12	0.98	0.02	0.44	0.56	0.91	0.09	0.57	0.43	0.92	0.08	0.35	0.65
13	0.98	0.02	0.44	0.56	0.91	0.09	0.57	0.43	0.92	0.08	0.35	0.65
14	0.98	0.02	0.44	0.56	0.91	0.09	0.57	0.43	0.92	0.08	0.35	0.65
15	0.98	0.02	0.44	0.56	0.91	0.09	0.57	0.43	0.92	0.08	0.35	0.65

This table reports variance decomposition of forecast errors from 1 to 15 period for bivariate VAR with $\xi_{SRP,t}$ and $\xi^{i}_{\widetilde{Bward},t}$ using sample from 26th June 2015 to 30th September 2015. The $\xi_{SRP,t}$ and $\xi^{IC}_{\widetilde{Bward},t}$ are obtained in the first-stage regression with specifications of $X_t^{SRP} = [r_t^{50\ ETF}, RV_t^{50\ ETF}]'$ and $X_t^i = [M_t, r_t^i, RV_t^i]'$ where $i \in \{IC,\ IF\ IH\}$, respectively. Panel 4a, Panel 4b and Panel 4c report results for bivariate VARs with $\xi_{SRP,t}$ and $\xi^{IC}_{\widetilde{Bward},t}$, with $\xi_{SRP,t}$ and $\xi^{IF}_{\widetilde{Bward},t}$, and with $\xi_{SRP,t}$ and $\xi^{IH}_{\widetilde{Bward},t}$, respectively. The second row lists forecasted variables, and the third row lists variables which are used for forecasting.

Table 5. Stage 1 regression results in extended sample.

	$\widetilde{Bward}_t^{IC}$	$\widetilde{Bward}_t^{IF}$	$\widetilde{Bward}_t^{IH}$	$\widetilde{Bward}_t^{IC}$	$\widetilde{Bward}_t^{IF}$	$\widetilde{Bward}_t^{IH}$	SRP
Intercept	− 2.51***	0.55***	0.91***	− 3.92***	− 0.70	− 1.28**	− 2.13**
	(0.30)	(0.24)	(0.18)	(0.54)	(0.47)	(0.62)	(1.16)
$Maturity_t$	− 0.06***	− 0.05***	− 0.03***	− 0.06***	− 0.04***	− 0.03***	
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
r_t^i	0.01	− 0.02	0.01	0.03	− 0.01	0.01	
	(0.04)	(0.06)	(0.05)	(0.04)	(0.06)	(0.06)	
RV_t^i	− 0.50***	− 0.73***	− 0.55***	− 0.57***	− 0.71***	− 0.60***	
	(0.13)	(0.28)	(0.19)	(0.11)	(0.27)	(0.20)	
$RS^i_{f,t}$				− 4.79	− 17.43***	− 6.33*	
				(3.67)	(6.62)	(3.57)	
$RS^i_{s,t}$				12.51	12.42	20.55***	
				(9.51)	(7.78)	(5.47)	
r_t^{50ETF}							− 0.34
							(0.39)
RV_t^{50ETF}							4.27***
							(1.37)
Adj. R^2	0.37	0.37	0.33	0.39	0.41	0.39	0.39

This table reports estimators of the stage 1 regression using the sample from 26th June 2015 to 29th December 2017. The regression is specified as follows: $\widetilde{Bward}_t^i = \alpha^i_{\widetilde{Bward}} + \beta^i_{\widetilde{Bward}} X_t^i + \xi^i_{\widetilde{Bward},t}$, $i \in \{IC, IF, IH\}$. For the second, third and fourth columns, $X_t^i = [M_t, r_t^i, RV_t^i]'$; for the fifth, sixth and seventh columns, $X_t^i = [M_t, r_t^i, RV_t^i, RS^i_{s,t}, RS^i_{f,t}]'$; for the last column, $X_t^{SRP} = [r_t^{ETF}, RV_t^{ETF}]'$. Standard deviations are reported in parentheses which are based on Newey–West corrections. *, ** and *** denote statistical significance at 10%, 5% and 1% levels, respectively.

In particular, this relationship is stronger during turmoil periods than it is during calm and normal periods. Therefore, the futures backwardation does reflect the risk premium paid by futures traders to hedge the tail risk when tail events occur.

4.5. Robustness tests

As the impulse response analysis of the VAR is order-sensitive, alternative endogenous variable order is used to test the robustness of the analysis. We also extend the sample

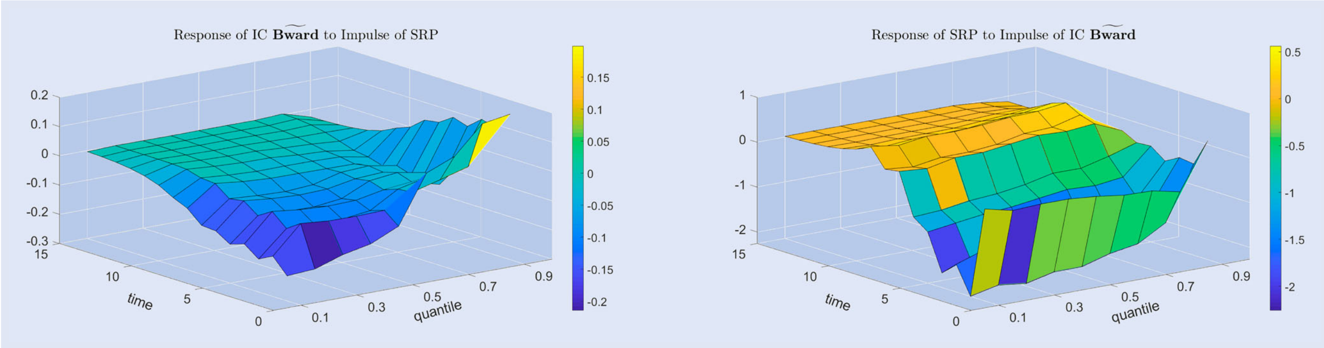

Figure 7. Impulse response functions for the bivariate MVMQ-CAViaR with SRP_t and $\widetilde{Bward}_t^{IC}$.
The Z-axis (perpendicular to the plane of the screen) represents impulse response. We use different colors to distinguish the level of the impulse response, lighter colors indicate more positive impact, while darker colors represent more negative impact.

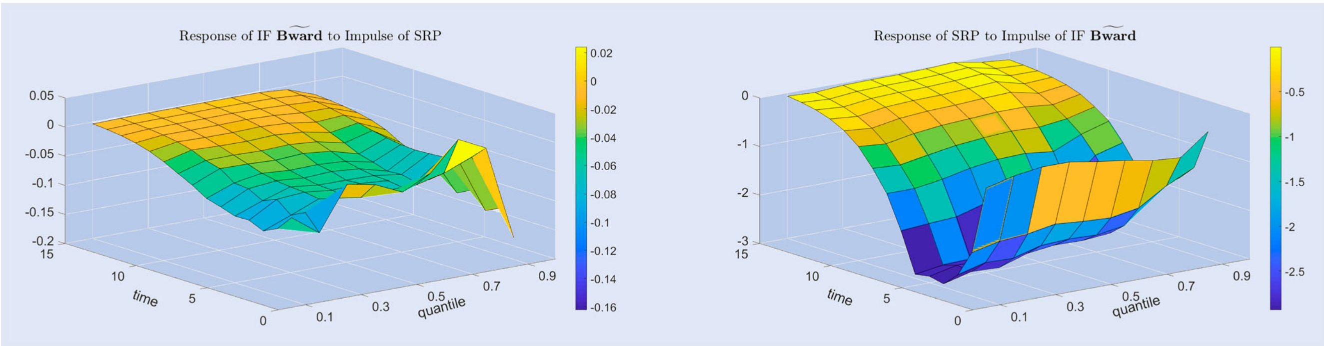

Figure 8. Impulse response functions for the bivariate MVMQ-CAViaR with SRP_t and $\widetilde{Bward}_t^{IF}$.
The model specification to obtain this figure is similar with figure 7.

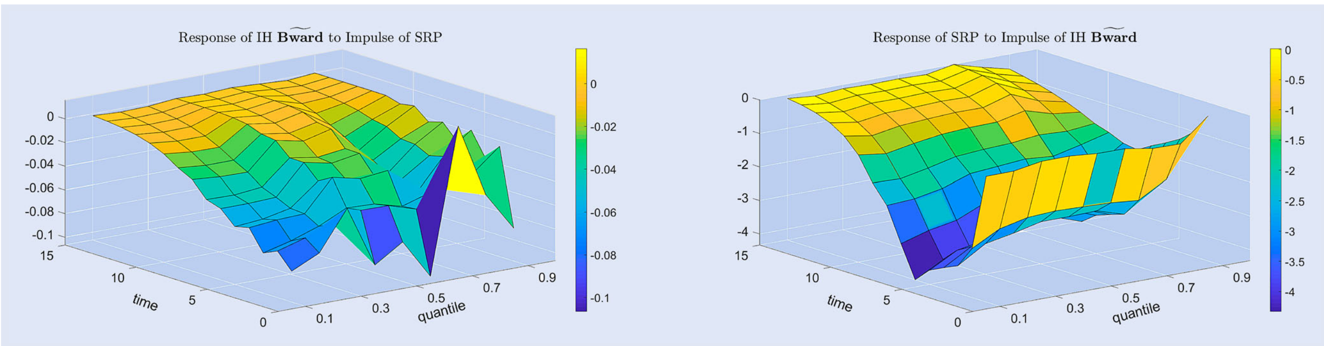

Figure 9. Impulse response functions for the bivariate MVMQ-CAViaR with SRP_t and $\widetilde{Bward}_t^{IH}$.
The model specification to obtain this figure is similar with figure 7.

period to 31st December 2015 and consider controlling for liquidity variables in the regressions. Owing to space constraints, these estimated results are reported in details in sections II.1, II.2 and II.3 of the appendix. Overall our main findings are quite robust to these variations.

Another concern is that the above empirical findings may be due to non-synchronous trading. As argued in section I and shown in the appendix, we have tried to mitigate the effect of non-synchronous trading by using the corresponding index ETF prices to restore the true level of the index and hence the true basis for each futures contract. However, some estimation errors may remain because of the no-arbitrage relationship between the ETFs and their corresponding indices. While it is difficult, if not impossible, to verify the validity of our method of isolating the non-synchronous trading effect, we are lucky enough to find that the period from 1st January 2020 to 31st July 2020 is very similar to our original sample, in the sense that the market also observes a high level of tail risk premium or fear resulted from the spread of COVID-19. Meanwhile, there is an extremely low number of firm stocks hitting price limits or getting suspended, indicating that the non-synchronous issue may be minimal during this period. Hence, we re-estimate our models using this new sample period. The results, reported in section II.4 of the appendix, are very similar to our main findings, and the conclusion about the relationship between

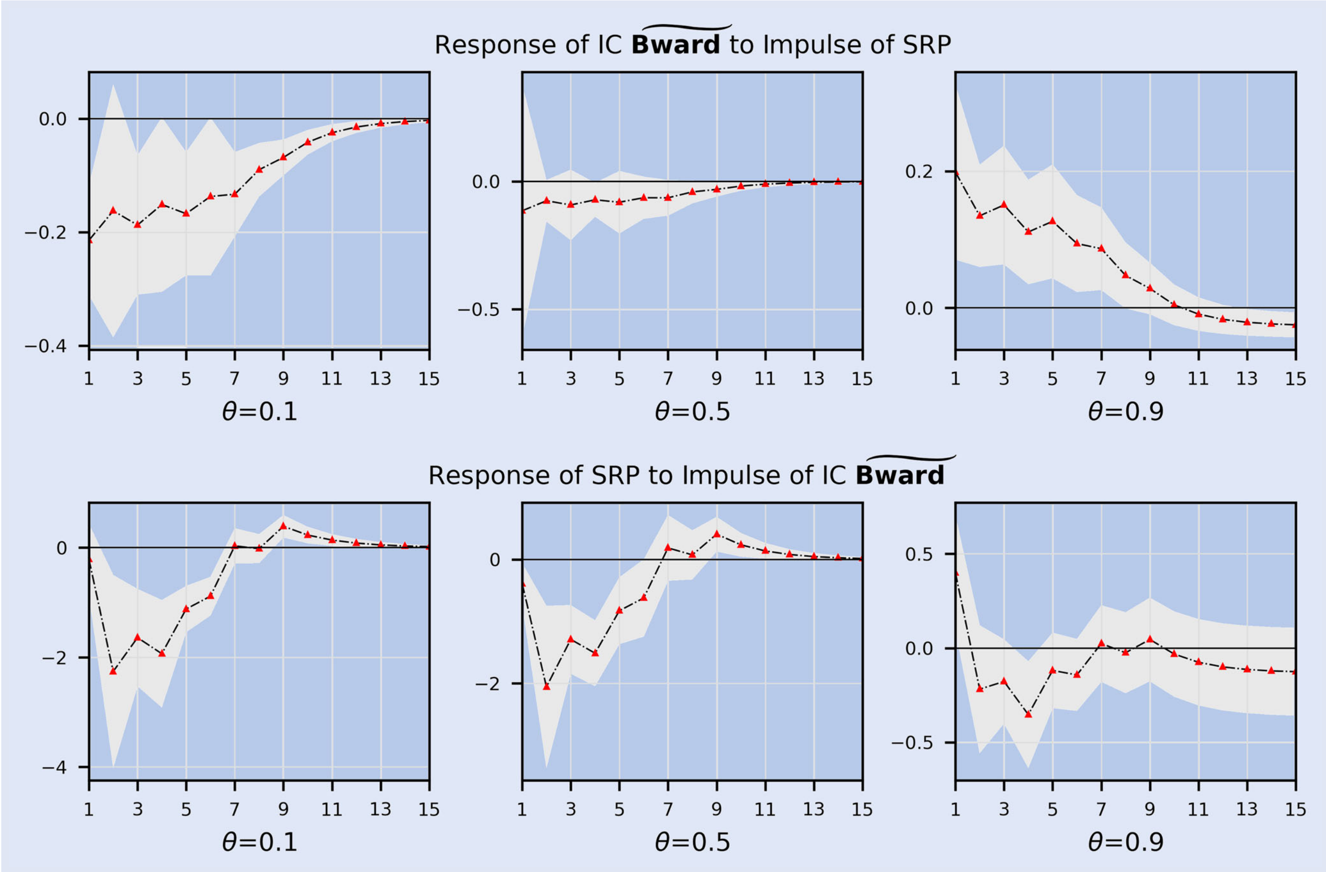

Figure 10. Impulse response functions for the bivariate MVMQ-CAViaR with SRP_t and $\widetilde{Bward}_t^{IC}$.
Impulse response functions of bivariate MVMQ-CAViaR with $\xi_{SRP,t}$ and $\xi^{IC}_{\widetilde{Bward},t}$ estimated in sample from 26th June 2015 to 29th December 2017 are presented. $\xi^{IC}_{\widetilde{Bward},t}$ and $\xi_{SRP,t}$ are obtained in the first-stage regression with specifications of $X_t^{SRP} = [r_t^{50\ ETF}, RV_t^{50\ ETF}]'$ and $X_t^{IC} = [M_t, r_t^{IC}, RV_t^{IC}, RS_{s,t}^{IC}, RS_{f,t}^{IC}]'$, respectively. The dynamic behavior of the response over time for different market states together with 95% confidence bands indicating the statistical significance of estimates are displayed.

the SRP and the backwardation of index futures does not change.

We also check whether other sentiment or uncertainty measures such as implied volatility index can be used to proxy the tail risk premium and hence hedging pressure. For this purpose, we substitute the SSE 50 ETF option implied volatility index, iVIX, for the SRP in our empirical tests. The results are reported in section II. 5 of the appendix. We have found no significant granger-causality relationships between $\xi_{iVIX,t}$ and $\xi^{i}_{\widetilde{Bward},t}$ where $i \in \{IC, IF, IH\}$, and no significant response of $\xi^{i}_{\widetilde{Bward},t}$ to the impulse of $\xi_{iVIX,t}$. These results suggest that the SRP measure contains information about market tail risk premium which is not captured by the model-free implied volatility. In fact, from the lower and right subfigure of figure 3, we can see that the time series of $\widetilde{Bward}$ has much higher variation than that of iVIX^2. In particular, during the period from 26th August to 7th September 2015 when the $\widetilde{Bward}$ hits new high, the difference between $\widetilde{Bward}$ and iVIX^2 is significant and apparent.

In summary, our results are quite robust to various issues including the ordering of endogenous variables, different sample periods, controlling for liquidity variables, and non-synchronous trading. We also find that the option implied volatility, iVIX, fails to explain the backwardation dynamics of Chinese index futures.

5. Further discussions

5.1. Why does the increased SRP decrease the futures basis?

The SRP, computed as the difference between the upside variance risk premium and the downside variance risk premium of investors, represents the 'net' pessimistic sentiment of the market and the uncertainty over market collapse (Bollerslev and Todorov 2011). The existing literature also confirms that investors' fear of uncertainty is mainly about rare events, e.g. large negative jumps (Liu *et al.* 2005) and skewness (Orlik and Veldkamp 2014), rather than the risk of the upside variance of the market. This coincides with the sub-sample interval (from 26th June to 7th September) covered in this study. The fear of tail risk triggers panic in the market, increases hedging pressure, further depresses the prices of index futures, thus leading to deep backwardation.

The most likely market mechanism is as follows: index futures has long been the only hedging tool in China

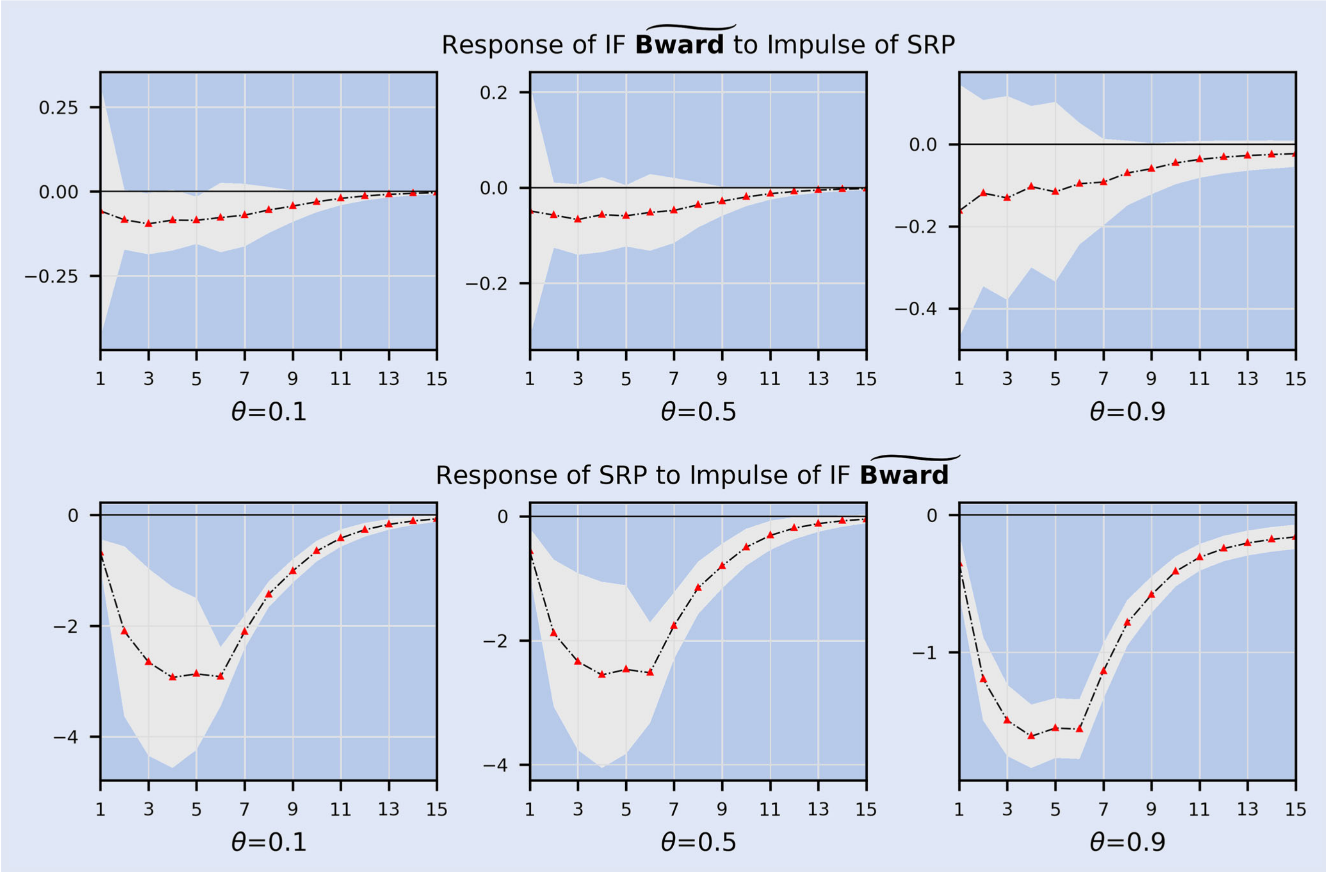

Figure 11. Impulse response functions for the bivariate MVMQ-CAViaR with SRP_t and $\widetilde{Bward}_t^{IF}$. The model specification to obtain this figure is similar with figure 10.

for institutional investors to hedge market-wide risk with reasonably good liquidity. This is true especially in midst of the market crash when the regulatory authorities launched restrictions such as prohibiting mutual funds from selling shares and short-sale constraints. When the Chinese stock market starts collapsing, huge hedging pressure emerges and is let out entirely on the index futures market. Speculators become less willing to long index futures in the vicinity of the current index level because they worry about further price decline. Therefore, bulls make an even lower bid for index futures contracts, whereas the hedging demand of bears accumulates quickly, leading to a growing imbalance between the supply and demand of hedging services and thereby causing long speculators to charge short hedgers a higher risk premium than usual. This risk premium represents the 'service fee' earned by speculators in providing hedging services to the market.

5.2. *Why do different contracts respond to the SRP differently?*

The above empirical results have found that the effects of tail risk premium on the prices of three futures contracts are in the decreasing order of IC, IF and IH, whereas the opposite effects of the backwardations of the three futures contracts on the tail risk premium are in the decreasing order of IH, IF and IC. We propose an explanation of this phenomenon below.

Kelly *et al.* (2016) argue that the market intervention and bailout measures taken by the U.S. government during the 2008 financial crises lower the put option prices of banks and indices as well as the skewness of the implied volatility. This argument applies to the Chinese case as well. The constituents of the underlying index of IH contracts are mainly large-cap companies, with banks and other financial companies accounting for approximately 60% of the total market capitalization. Most of these large-cap stocks are also included in the underlying index of IF contracts. During the stock market crash, these shares tend to have high bailout priority. In addition, just as in the U.S. market, financial institutions are 'too systemic to fail' (Kelly *et al.* 2016). In contrast, the constituent stocks for IC contracts are mainly small non-financial companies and are 'less systemic', thus having lower bailout priority. As a result, investors think that holding shares of large companies is akin to having an 'insurance', and are thus less worried about the tail risk of the index that has constituent shares of large companies than they do about the tail risk of other indices. Therefore, IH and IF contracts, the former in particular, are less sensitive to investors' tail risk aversion, whereas IC contracts are more sensitive.

5.3. *How can the sustained backwardation be eliminated?*

As discussed in Introduction, the futures backwardation is a trading cost for hedgers. A 3% monthly backwardation, as

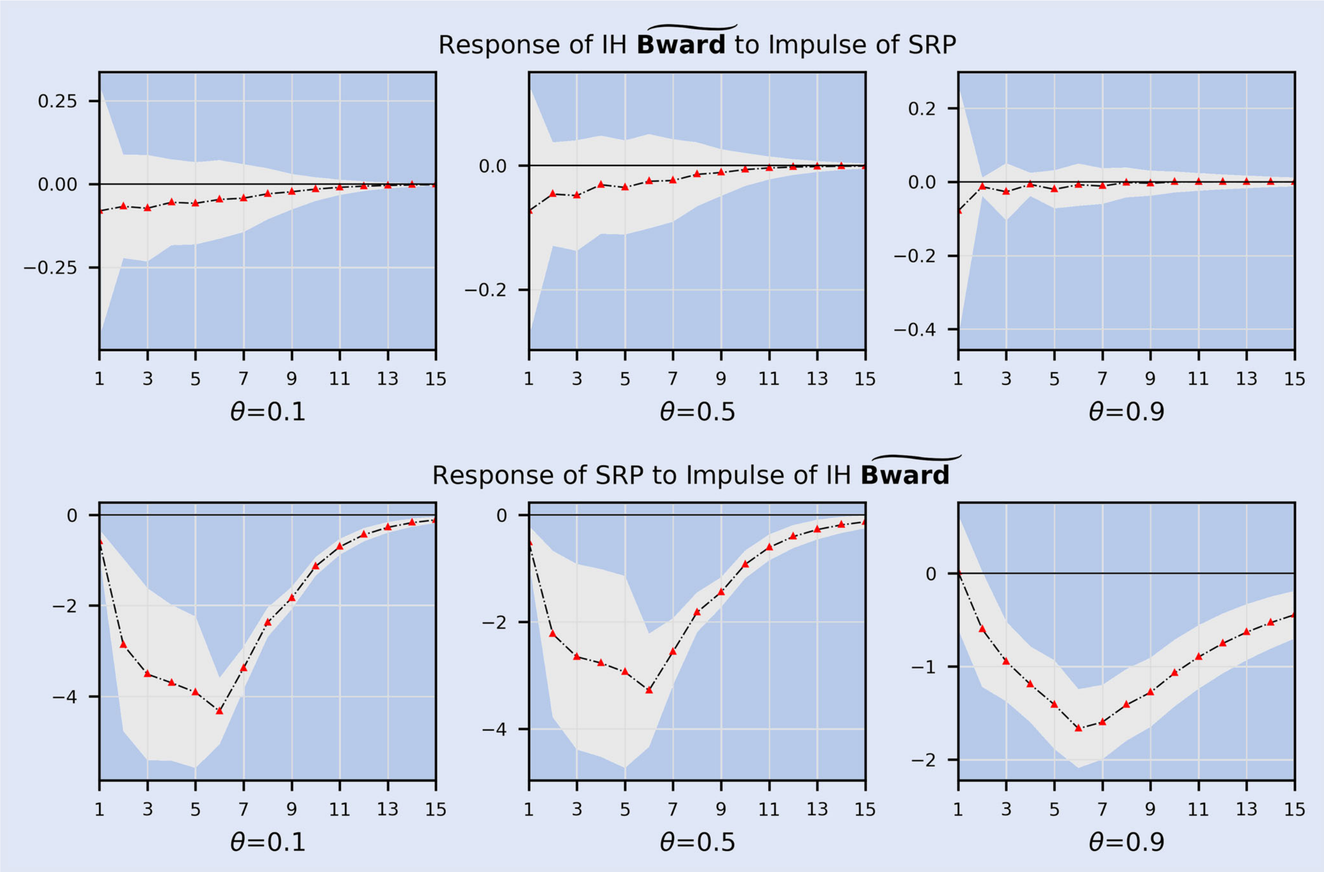

Figure 12. Impulse response functions for the bivariate MVMQ-CAViaR with SRP_t and $\widetilde{Bward}_t^{IH}$. The model specification to obtain this figure is similar with figure 10.

one observes during the 2015 Chinese market crash, would imply an annualized hedging cost of 36%, which is unbelievably high even from a global perspective. Hence, getting the futures backwardation back to its normal level becomes an urgent agenda for the regulators.

So, will a regulatory ban on index futures trading help to reduce the severity of backwardations? The answer to this question, based on our study, is clearly no.

Futures is an essential risk management tool. During a liquidity collapse such as the one in the study, index futures provide an alternative to short selling spot. When futures basis is pushed sufficiently low due to hedging pressures, short stocks and long futures can generate risk-free arbitrage profits. The entry of arbitrageurs into the market will fix and drive the basis back to normal, thus reducing the tail risk aversion of market investors. However, the regulatory short sale constraints in the Chinese A-share market,† together with the futures trading limits, have made it impossible to effectively execute arbitrage transactions. As shown in figure 13, the outstanding value of open index futures positions is much larger than the total securities loan balance of their corresponding constituent stocks. The value of outstanding IC contracts is over 100 times, even over 400 times at one point, that of the total securities loan balance of the corresponding index-constituent stocks; for IF and IH contracts, the numbers are smaller but remain at over 10 times. Because the securities loan market capacity is much lower than that of the index futures market, it is impossible to fix the deep discounts in the index futures markets through arbitrage transactions, leading to persistent futures backwardations.

Not surprisingly, it turns out that these trading restrictions on the spot and index futures markets are not only unable to eliminate the sustained negative basis of index futures contract prices but also detrimental to the risk management function of the index futures markets by increasing investors' hedging costs. One piece of advice our analysis provides is that regulators should enlarge the toolset of financial risk management in China by launching more derivatives and relaxing the short sale constraints, in the hope to meet investors' hedging demand and removing the origin of the tail risk aversion and panic emotions that are transmitted to the spot market from the index futures market.

† According to Cao and Li (2020), China's MTSL (margin trading and security lending) remains unbalanced with MT: SL balance at 97:3, mainly due to limited eligible stocks/lenders/order types; and high fees for stock lending. The industry stock lending rate is around 8%, a bit higher than the margin financing rate. In addition, security lending business in China is rather restricted, both because brokers face large hedging costs for their security pools so that there is a limited supply of securities, and because there are substantial communication/negotiation costs between brokers and typical clients. All of these short-sale constrains help generate a wide no-arbitrage interval on index futures bases, resulting in the 'persistency' of deep backwardation.

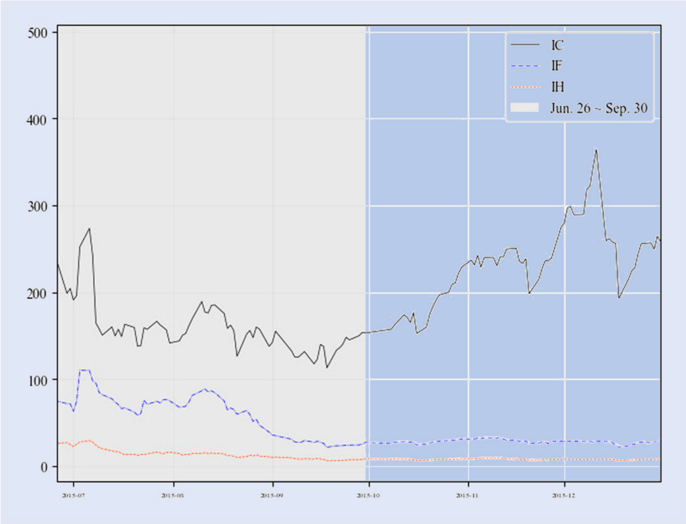

Figure 13. Ratio of volume and open interest of index futures. The greyed regions correspond to the period from 26th June 2015 to 30th December 2015.

5.4. *Why did the sustained negative basis start when the stock market crashed?*

The documented sustained backwardation of index futures is a new phenomenon. It started on 26th June 2015, and before that date it is uncommon to find a sustained negative basis in the Chinese index futures market, or elsewhere in the global index futures market. As discussed above, such a sustained large deviation in spot and futures prices is clearly associated with limits to arbitrage. However, since market frictions such as short sale constraints have long been present in the Chinese stock market, it is difficult to use the market friction-induced arbitrage limitation to explain why this phenomenon emerged at the beginning of the market collapse.

The following questions naturally arise: does it have a cause similar to the emergence of the option implied volatility smile in the 1987 US market crash, i.e. a 'peso effect,' or can it be explained by the market sentiment hypothesis proposed in Lee *et al.* (1991) to solve the puzzle of closed-end fund discount? Unfortunately, given that the data available is short, we are unable to present a complete and clear answer to these important questions at this stage.

6. Conclusion

In this study, we examine the causes of the persistent and significant backwardation of Chinese index futures during the 2015 market crash. We first correct the distorted index caused by non-synchronous trading in the spot and futures markets. The 'true' basis uncovered reveals that adjustment of such infrequent trading factors as trade suspensions and daily price limits is insufficient to explain the sustained backwardation in the index futures markets. Attempts to use the spot market volatility and return to explain the changes of basis as in Chen *et al.* (1995) also fail.

We then compute the skewness risk premium embodied in the SSE 50 ETF option prices as a proxy of investors' level of net pessimism, panic, and the uncertainty premium paid for tail events. The VAR analysis on the SRP and the bases of index futures in sample during market crash indicates that the negative bases of index futures can be well explained by the SRP, which is related to traders' panic triggered by the fear of tail risks. Speculators' fear of the crash risk leads them to charge high 'insurance premium' on hedgers. On the other hand, the deepening backwardation of index futures accompanies an increasing SRP, leading to a new round of investors' panic and an even higher premium charged on market uncertainty. The MVMQ-CAViaR analysis further corroborates these findings from the VAR analysis. Thus our results supplement the existing literature by showing that hedging pressure does impact futures prices and vice versa.

The deep backwardation of index futures persists throughout the market crash because of the short sale constraints in the spot market. In the presence of short sale constraint, it is technically impossible for arbitrageurs to enter the markets to eliminate the deep backwardation.† Regulators should strive to improve the market microstructure by increasing the supply of risk management 'tool and services' and reducing the uncertainty perceived by investors, thereby lowering hedging costs.

Results in this paper need to be cautiously interpreted though, because we provide at most indirect evidence. However, direct evidence is hard to obtain at this moment as the futures positions data released by the CFFEX are the aggregate gross client positions of futures commissioners rather than the categorized client positions.‡ In the future, if the categorized client positions data becomes available, then the mechanism described above can be rigorously tested.

Acknowledgments

We thank two anonymous Quantitative Finance referees, for constructive suggestions that significantly improved the paper. We also thanks the participants of the Xiamen University-Tianjin University Forum on Quantitative Finance and Risk Management, the 1st Greater China Finance Symposium at Xiamen University, and the annual China Finance Research Conference at Tsinghua University, where earlier versions of this paper were presented. We have benefited from comments and suggestions by Qi Liang, Wei Zhang, Zongwu Cai, Yinggang Zhou, Xiaoyan Zhang, Qunzi Zhang. We are solely responsible for any remaining errors and omissions.

Disclosure statement

No potential conflict of interest was reported by the author(s).

† Anecdote has it that a few market investors replicate short positions from option portfolios, but soon the regulatory body forbids this kind of behavior.

‡ The CFTC classifies traders of commodity futures into business positions and speculative positions, and the former can be grouped into hedging positions.

Funding

Liang's research is supported by Natural Science Foundation of China (Grant No. 71871087 and 71971077), Humanities and Social Science Fund of Ministry of Education of China (Grant No. 17YJC790089) and Natural Science Foundation of Hunan Province, China (Grant No. 2019JJ40026). Han's research is supported by Natural Science Foundation of China (NSFC project no. 71471153).

Supplemental data

Supplemental data for this article can be accessed online at http://dx.doi.org/10.1080/14697688.2024.2330612.

ORCID

Jufang Liang http://orcid.org/0000-0003-0917-8799

References

Acharya, V.V., Lochstoer, L.A. and Ramadorai, T., Limits to arbitrage and hedging: Evidence from commodity markets. *J. Financ. Econ.*, 2013, **109**, 441–465.
Antoniou, A. and Garrett, I., To what extent did stock index futures contribute to the October 1987 stock market crash? *Econ. J.*, 1993, **103**, 1444–1461.
Bates, D.S., The market for crash risk. *J. Econ. Dyn. Control*, 2008, **32**, 2291–2321.
Bessembinder, H., Systematic risk, hedging pressure, and risk premiums in futures markets. *Rev. Financ. Stud.*, 1992, **5**, 637–667.
Bollen, N.P. and Whaley, R.E., Does net buying pressure affect the shape of implied volatility functions? *J. Finance*, 2004, **59**, 711–753.
Bollerslev, T. and Todorov, V., Tails, fears, and risk premia. *J. Finance*, 2011, **66**, 2165–2211.
Bollerslev, T., Gibson, M. and Zhou, H., Dynamic estimation of volatility risk premia and investor risk aversion from option-implied and realized volatilities. *J. Econom.*, 2011, **160**, 235–245.
Cao, H. and Li, Y., *China Securities Sector—A Securities Lending a Blue Ocean Opportunity*, 2020 (UBS).
Carr, P. and Wu, L., Variance risk premiums. *Rev. Financ. Stud.*, 2009, **22**, 1311–1341.
Chen, N.F., Cuny, C.J. and Haugen, R.A., Stock volatility and the levels of the basis and open interest in futures contracts. *J. Finance*, 1995, **50**, 281–300.
De Roon, F.A., Nijman, T.E. and Veld, C., Hedging pressure effects in futures markets. *J. Finance*, 2000, **55**, 1437–1456.
Drechsler, I., Uncertainty, time-varying fear, and asset prices. *J. Finance*, 2013, **68**, 1843–1889.
Drechsler, I. and Yaron, A., What's vol got to do with it. *Rev. Financ. Stud.*, 2011, **24**, 1–45.
Feunou, B., Jahan-Parvar, M.R. and Okou, C., Downside variance risk premium. *J. Financ. Econom.*, 2018, **16**, 341–383.
Garleanu, N., Pedersen, L.H. and Poteshman, A.M., Demand-based option pricing. *Rev. Financ. Stud.*, 2008, **22**, 4259–4299.
Han, J. and Pan, Z., On the relation between liquidity and the futures-cash basis: Evidence from a natural experiment. *J. Financ. Mark.*, 2017, **36**, 115–131.
Harris, L., The October 1987 S&P 500 stock-futures basis. *J. Finance*, 1989, **44**, 77–99.
Hemler, M.L. and Longstaff, F.A., General equilibrium stock index futures prices: Theory and empirical evidence. *J. Financ. Quant. Anal.*, 1991, **26**, 287–308.
Hicks, J. R., *Value and Capital*, 1946 (Oxford University Press: Oxford).
Hirshleifer, D., Determinants of hedging and risk premia in commodity futures markets. *J. Financ. Quant. Anal.*, 1989, **24**, 313–331.
Kelly, B., Lustig, H. and Van Nieuwerburgh, S., Too-systemic-to-fail: What option markets imply about sector-wide government guarantees. *Am. Econ. Rev.*, 2016, **106**, 1278–1319.
Keynes, J. M., *A Treatise on Money*, Vol II, 1930 (MacMillan: London).
Lee, C.M., Shleifer, A. and Thaler, R.H., Investor sentiment and the closed-end fund puzzle. *J. Finance*, 1991, **46**, 75–109.
Liu, J., Pan, J. and Wang, T., An equilibrium model of rare-event premia and its implication for option smirks. *Rev. Financ. Stud.*, 2005, **18**, 131–164.
MacKinlay, A.C. and Ramaswamy, K., Index-futures arbitrage and the behavior of stock index futures prices. *Rev. Financ. Stud.*, 1988, **1**, 137–158.
Miao, J., Wei, B. and Zhou, H., Ambiguity aversion and the variance premium. *Q. J. Finance*, 2019, **9**, 1950003.
Orlik, A. and Veldkamp, L., Understanding uncertainty shocks and the role of black swans. Technical report, National Bureau of Economic Research, 2014.
Roll, R., Schwartz, E. and Subrahmanyam, A., Liquidity and the law of one price: The case of the futures-cash basis. *J. Finance*, 2007, **62**, 2201–2234.
White, H., Kim, T.H. and Manganelli, S., VAR for VaR: Measuring tail dependence using multivariate regression quantiles. *J. Econom.*, 2015, **187**, 169–188.