

Wang, Xinyu; Zhang, Lele; Cheng, Qiuying; Shi, Song; Niu, Huawei

## Article

# What drives risk in China's soybean futures market? Evidence from a flexible GARCH-MIDAS model

Journal of Applied Economics

## Provided in Cooperation with:

University of CEMA, Buenos Aires

*Suggested Citation:* Wang, Xinyu; Zhang, Lele; Cheng, Qiuying; Shi, Song; Niu, Huawei (2022) : What drives risk in China's soybean futures market? Evidence from a flexible GARCH-MIDAS model, Journal of Applied Economics, ISSN 1667-6726, Taylor & Francis, Abingdon, Vol. 25, Iss. 1, pp. 454-475,  
<https://doi.org/10.1080/15140326.2022.2046989>

This Version is available at:

<https://hdl.handle.net/10419/314171>

## Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

## Terms of use:

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*



<https://creativecommons.org/licenses/by/4.0/>



## What drives risk in China's soybean futures market? Evidence from a flexible GARCH-MIDAS model

Xinyu Wang, Lele Zhang, Qiuying Cheng, Song Shi & Huawei Niu

**To cite this article:** Xinyu Wang, Lele Zhang, Qiuying Cheng, Song Shi & Huawei Niu (2022) What drives risk in China's soybean futures market? Evidence from a flexible GARCH-MIDAS model, *Journal of Applied Economics*, 25:1, 454-475, DOI: [10.1080/15140326.2022.2046989](https://doi.org/10.1080/15140326.2022.2046989)

**To link to this article:** <https://doi.org/10.1080/15140326.2022.2046989>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.



Published online: 05 Apr 2022.



Submit your article to this journal [↗](#)



Article views: 2273



View related articles [↗](#)



View Crossmark data [↗](#)



Citing articles: 6 View citing articles [↗](#)

# What drives risk in China's soybean futures market? Evidence from a flexible GARCH-MIDAS model

Xinyu Wang<sup>a</sup>, Lele Zhang<sup>b</sup>, Qiuying Cheng<sup>a</sup>, Song Shi<sup>a</sup> and Huawei Niu<sup>a</sup>

<sup>a</sup>School of Economics and Management, China University of Mining and Technology, Xuzhou, Jiangsu, China; <sup>b</sup>School of Business and Management, Shanghai International Studies University, Shanghai, China

## ABSTRACT

Modeling futures market risk simultaneously influenced by macro low-frequency information and daily risk factors is a valuable challenge. We propose a new general framework for it based on the flexible GARCH-MIDAS model. It uses a skewed t distribution to describe the asymmetry of long and short trading positions, allows for a different number of trading days per month, and can identify the optimal combination of risky factors. We also derive its impact response function on how low-frequency factors directly influence the high-frequency futures market risk. Through an exhaustive empirical analysis of the Chinese soybean futures market, we not only find its excellent out-of-sample market risk forecasting performance but also offer systematic recommendations for improving risk management.

## ARTICLE HISTORY

Received 22 August 2021  
Accepted 19 February 2022




## KEYWORDS

GARCH-MIDAS; volatility; value at risk; skewed t distribution; futures market

## 1. Introduction

Due to the influence of financialisation, since the mid-2000s, almost commodity markets have witnessed turbulent time periods and become more vulnerable to unconventional shocks such as market sentiments, policy uncertainties, and other unexpected events (Zhang & Ji, 2019). Price volatility has a significant impact on production and investment decisions. Producers and consumers face increased price uncertainty and may suffer higher borrowing costs and greater cash flow volatility. In practice, advanced risk measurement models are needed to manage risk and avoid losses.

As the largest emerging economy, China ranks among the top commodity traders in the world. Among agricultural commodities, soybean products are of significant importance since they represent a major component of household consumption and have a critical influence on food security (Ordu, Oran, & Soytaş, 2018). Compared to other international soybean futures markets, China's soybean futures market has its own peculiarities. According to the China Futures Association (CFA), China imports more than eighty percent of its total soybean consumption, which reached 85.511 million tons in 2019, accounting for 60% of global soybean trade. Given the huge domestic demand, the Ministry of Agriculture and Rural Affairs of China has decided to implement the

**CONTACT** Xinyu Wang  [wangxinyu@cumt.edu.cn](mailto:wangxinyu@cumt.edu.cn)  School of Economics and Management, China University of Mining and Technology, Xuzhou, Jiangsu, China; Lele Zhang  [zhanglele@shisu.edu.cn](mailto:zhanglele@shisu.edu.cn)  School of Business and Management, Shanghai International Studies University, Shanghai, China

© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

soybean revitalization plan since 2019. The domestic soybean planting area reaches 140 million acres in 2020 to maintain a safe supply. In particular, due to the dietary habits of the Chinese people, there is a high demand for non-GMO soybeans – because Chinese food regulation requires that soybean oil for food can only be pressed from non-GMO soybeans and that GMO soybean oil cannot be circulated in the market. To ensure food supply safety, China is expanding the cultivation and import of non-GMO soybeans. Recently, the uncertain trading relationship and policy adjustments have caused sharp volatility of China soybean futures price. In the past, Chinese soybeans were mainly sourced from the US, but trade tensions between the US and China have forced China to open up wider international trade channels since July 2018 (Hua & Chen, 2007; Jia, Wang, Tu, & Li, 2016; Liu, Wang, Wang, Wu, & Zhang, 2015; Zhao, Yang, Zhang, & Qi, 2010).

The soybean futures market plays an important role in stabilizing spot prices because of the price discovery function. The risk of price fluctuations in the spot market can be hedged through soybean futures trading. Managing the soybean futures market risk well is crucial for Chinese investors to avoid huge loss. National urgent needs prompted us to delve into the development of advanced risk measurement models. Because of the common characteristics of commodity futures markets, our models can also provide a reference for studying other markets. We prefer to use an analytical approach to estimate market risk based on a volatility model. About risk metrics, we use the value at risk (VaR) commonly used in the financial industry to measure the level of risk exposure, which calculates the maximum loss expected (or worst case scenario) on an investment with a given probability over a pre-specified horizon.

On selecting proper risk models, we decide to adopt the mixed-frequency modelling approach, because commodity futures prices are measured at a relatively high frequency of data, but considerable risky factors of commodity futures are monthly macroeconomic variables, i.e., daily. Ghysels, Sinko, and Valkanov (2007) propose a popular mixed data sampling (MIDAS) approach to solve it. Further, Engle, Ghysels, and Sohn (2013) propose a GARCH-MIDAS model that decomposes the conditional variance into the product of long-term and short-term components of volatility. Several studies have explored the drivers of commodity volatility using the GARCH-MIDAS with a normal distribution. For instance, Donmez and Magrini (2013) find the monetary factors are essential to describe agricultural product price volatility, while Mo, Gupta, Li, and Singh (2018) suggest both domestic and international macroeconomic information plays an important role in determining the price volatility of emerging commodity futures. The GARCH-MIDAS model is also effective in predicting the volatility of high-frequency financial asset returns using macro variables (Asgharian, Hou, & Javed, 2013; Conrad & Kleen, 2019; Conrad, Loch, & Rittler, 2014; Fang, Chen, Yu, & Qian, 2018). Theodossiou (1998) proposes a skewed *t* distribution to describe the asymmetry of financial assets returns. In contrast to the GARCH-MIDAS model (Engle et al., 2013), Wang, Luo, Wang, Xu, and Wu (2021) propose a single-factor GARCH-MIDAS model that uses a skewed *t* distribution (GARCH-MIDAS-Skewed *t*).

Considering the complex scenario of China's Soybean Futures Market, we contribute to develop a new systematical framework that can measure market risk more accurately than before, the core of which is a flexible GARCH-MIDAS model. Regarding to the heterogeneous behavior of long and short positions, we propose a multi-factor GARCH-

MIDAS model using a skewed t-distribution. We construct a new complete risky factor set from the aspects of supply and demand, substitutes, downstream products, related financial markets, and economic conditions. Then, we successfully identify the best factor combination by eliminating the strong correlations among factors and comparing their significances by variance ratios. More importantly, we derive for the first time a dynamic equation on how low-frequency variables directly affect the volatility or VaR of high-frequency variables. It can help us understand how macro factors affect daily risk. In addition, our model is computationally flexible, allowing for a different number of trading days per month in practice. Finally, by the backtesting of out-of-sample VaR predictions, our model proves to be more effective than the benchmark model.

The remainder of this paper is organized as follows: [Section 2](#) presents the details of the multi-factor GARCH-MIDAS-Skewed t model; [Section 3](#) performs an empirical analysis; [Section 4](#) gives the conclusion.

## 2. The new measurement framework for market risk

The classical GARCH-MIDAS (Engle et al., 2013) model is too simple, such as the assumption of normal distribution, to be suitable for analyzing futures markets. Therefore, from the practice of risk management in futures markets, we systematically propose a new risk measurement framework to improve the performance.

### 2.1. The flexible GARCH-MIDAS model

First, our model supports unbalanced mixed-frequency data structures, which are very common in practice but rarely discussed in the literature, that is, it allows for the various numbers of trading days per month, because holidays are unevenly distributed throughout twelve months in China. We denote the logarithmical return of the soybean future on day  $i$  in month  $t$  as  $r_{i,t}$ , where  $t = 1, \dots, T$  and  $i = 1, \dots, n_t$ .  $n_t$  is the number of days in month  $t$ . Because holidays are unevenly distributed throughout twelve months in China, our model allows for the various numbers of trading days ( $n_t$ ) per month.  $r_{i,t}$  obeys the following process:

$$r_{i,t} = \mu + \sqrt{\tau_t g_{i,t}} \varepsilon_{i,t}, \varepsilon_{i,t} | \Psi_{i-1,t} \sim F(0, 1), \quad (1)$$

where  $\Psi_{i-1,t}$  is the information set up to day  $i - 1$  of month  $t$ . The conditional variance  $\sigma_{i,t}^2$  is the product of the short-term component of volatility  $g_{i,t}$  on day  $i$  in month  $t$  and the long-term component of volatility  $\tau_t$  in month  $t$ . Like Wang et al. (2021), we also take the standardized skewed t distribution with zero mean and unit variance (Hansen, 1994) for the error item  $\varepsilon_{i,t}$ , which is denoted as  $F(0, 1)$ . Its density function is expressed as follows:

$$f(\varepsilon | \eta, \lambda) = \begin{cases} bc \left( 1 + \frac{1}{\eta-2} \left( \frac{b\varepsilon+a}{1-\lambda} \right)^2 \right)^{-\frac{\eta+1}{2}}, & \varepsilon < -a/b, \\ bc \left( 1 + \frac{1}{\eta-2} \left( \frac{b\varepsilon+a}{1+\lambda} \right)^2 \right)^{-\frac{\eta+1}{2}}, & \varepsilon \geq -a/b, \end{cases} \quad (2)$$

where  $\eta$  and  $\lambda$  are the degrees of freedom and skewness, respectively,  $2 < \eta < \infty$ , and  $-1 < \lambda < 1$ . The constants  $a$ ,  $b$ , and  $c$  are given by

$$\begin{cases} a = 4\lambda c \left( \frac{\eta-2}{\eta-1} \right), \\ b^2 = 1 + 3\lambda^2 - a^2, \\ c = \frac{\Gamma\left(\frac{\eta+1}{2}\right)}{\sqrt{\pi(\eta-2)\Gamma\left(\frac{\eta}{2}\right)}}. \end{cases} \quad (3)$$

*Second*, considering that the volatility on the first day of the month, different from other days, is influenced by macro factors of the previous month in a GARCH (1,1) process, we put forth an exact formula to calculate the short-term volatility dynamics, which is not adequately stated in (Engle et al., 2013), that is, if  $i = 1$ , the short-term component of volatility on the first day of the month will depend on the squared residual on the last day of the previous month as follows:

$$g_{1,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{n_{t-1},t-1} - \mu)^2}{\tau_{t-1}} + \beta g_{n_{t-1},t-1}. \quad (4)$$

However, starting from the second day in a month, that is, if  $i > 1$ , the short-term component of volatility is given by

$$g_{i,t} = (1 - \alpha - \beta) + \alpha \frac{(r_{i-1,t} - \mu)^2}{\tau_t} + \beta g_{i-1,t}, \quad (5)$$

where  $\alpha + \beta < 1$  holds in Eqs. (4) and (5).

*Third*, it is important to analyze the combined impacts of multiple factors on commodity futures volatility in the framework of GARCH-MIDAS. However, to our knowledge, there is no literature so far that discusses how to determine the optimal combination of factors. We propose a method to determine the optimal combination of factors. We first determine the contributions of risky factors that are measured by the variance ratio (Engle et al., 2013), that is,  $\text{VaR}(\tau_t)/\text{VaR}(\tau_t g_{it})$ . Then, we use stepwise regression to add them to the equation in order of importance from highest to lowest, and if the newly added factor is significant and can improve the variance contribution, this factor will be retained, and so on. For the set of variables with high correlations large than 0.5, we first select the one with the highest variance contribution into the equation, and if it is retained, the other variables in this group are eliminated; if this variable is eliminated, the variable with the next highest variance contribution is added, and the process is repeated until the end.

For instance, we introduce one low-frequency exogenous factor  $Z_t$  and one high-frequency exogenous factor  $x_{it}$  measured by logarithmic returns into  $\tau_t$  in turn as follows:

$$\tau_t = \exp(m + \theta_1 \sum_{k=1}^K \varphi_k(\omega_1) Z_{t-k}^{l(v)} + \theta_2 \sum_{k=1}^K \varphi_k(\omega_2) RV_{t-k}), \quad (6)$$

where  $Z_{t-k}^{l(v)}$  are the level and volatility of the  $k$ -order lag exogenous variables respectively.  $RV_t$  is the monthly realized volatility that equals  $\sum_{i=1}^{n_t} x_{i,t}^2$ .

In Eq. (6),  $Z_{t-k}^{l(v)}$  and  $RV_t$  are smoothed through the MIDAS regression. The single-parameter beta weight function is defined in (Engle et al., 2013), where we give the most recent information a larger weight,

$$\varphi_k(\omega) = \frac{(1 - k/K)^{\omega-1}}{\sum_{j=1}^K (1 - j/K)^{\omega-1}}. \quad (7)$$

If the coefficient  $\theta_2$  is significant and the variance contribution is increased after adding  $x_{it}$ , we will keep it. Other variables that are highly correlated with  $x_{it}$  if present will be eliminated.

Our model can use the likelihood-based approach to estimate the parameters. In the case of a single-factor model, we denote the parameter vector as  $\Theta = \{\mu, \alpha, \beta, m, \theta, \omega, g_0, \eta, \lambda\}$ .  $g_0$  is the initial parameter in the recursive Eq. (5). The log-likelihood function is presented as follows:

$$\text{LLF}(\Theta) = \sum_{i=1}^{n_t} \sum_{t=K+1}^T \log \left( \frac{1}{\sqrt{\tau_t(\Theta)g_{i,t}(\Theta)}} f \left( \frac{r_{i,t} - \mu}{\sqrt{\tau_t(\Theta)g_{i,t}(\Theta)}} \right) \right), \quad (8)$$

where  $f(\cdot)$  is the standardized skewed t density function in Eq. (2). We obtain the optimal estimated parameters by minimizing the negative LLF. The optimal lag order  $K$  is determined by the Bayesian information criterion (BIC).

## 2.2. The impact response function

*Fourth*, it is important to discover the direct responses of high-frequency volatility and VaR to low-frequency information, also called the impact response function. In the original GARCH-MIDAS model, the daily volatility is decomposed into the product of the short-term volatility component and the long-term volatility component, so it cannot directly reveal how low-frequency variables affect daily volatility or VaR. Here, we present for the first time the response functions of daily volatility and VaR to the impacts of low-frequency information. Without loss of generality, we take single-factor model to derive a direct expression of  $\sigma_{it}$ , that is,

$$\tau_t = \exp(m + \theta \sum_{k=1}^K \varphi_k(\omega) Z_{t-k}^{l(v)}). \quad (9)$$

Let  $\gamma = (1 - \alpha - \beta)\exp(m)$  and  $\Delta Z_{t-k}^{l(v)} = Z_{t-k}^{l(v)} - Z_{t-k-1}^{l(v)}$ . We substitute Eq. (9) in to Eq. (4), and then we have the conditional volatility  $\sigma_{it}^2$  as follows:

$$\sigma_{1t}^2 = \tau_t g_{1t} = (1 - \alpha - \beta)\tau_t + \frac{\tau_t}{\tau_{t-1}} (\alpha(r_{n_{t-1},t-1} - \mu)^2 + \beta\tau_{t-1}g_{n_{t-1},t-1}) \quad (10)$$

$$= \gamma \exp(\theta \sum_{k=1}^K \varphi_k(\omega) Z_{t-k}^{l(v)}) + \exp(\theta \sum_{k=1}^K \varphi_k(\omega) \Delta Z_{t-k}^{l(v)}) (\alpha(r_{n_{t-1},t-1} - \mu)^2 + \beta\sigma_{n_{t-1},t-1}^2). \quad (11)$$

We substitute Eq. (9) in to Eq. (5), and then we have the conditional volatility  $\sigma_{it}^2$  for  $i > 1$  as follows:

$$\sigma_{it}^2 = \tau_t g_{it} = (1 - \alpha - \beta)\tau_t + \alpha(r_{i-1,t} - \mu)^2 + \beta\tau_t g_{i-1,t} \quad (12)$$

$$= \gamma \exp\left(\theta \sum_{k=1}^K \varphi_k(\omega) Z_{t-k}^{l(v)}\right) + \alpha(r_{i-1,t} - \mu)^2 + \beta \sigma_{i-1,t}^2. \quad (13)$$

Therefore, we derive a reparameterized GARCH-MIDAS model based on Eqs. (1), (10), and (12) as follows:

$$\begin{cases} r_{i,t} = \mu + \sigma_{it} \varepsilon_{i,t}, \varepsilon_{i,t} | \Psi_{i-1,t} \sim F(0, 1), \\ \sigma_{1t}^2 = \gamma \exp\left(\theta \sum_{k=1}^K \varphi_k(\omega) Z_{t-k}^{l(v)}\right) + \exp\left(\theta \sum_{k=1}^K \varphi_k(\omega) \Delta Z_{t-k}^{l(v)}\right) (\alpha(r_{n_{t-1},t-1} - \mu)^2 + \beta \sigma_{n_{t-1},t-1}^2), \\ \sigma_{it}^2 = \gamma \exp\left(\theta \sum_{k=1}^K \varphi_k(\omega) Z_{t-k}^{l(v)}\right) + \alpha(r_{i-1,t} - \mu)^2 + \beta \sigma_{i-1,t}^2, \quad i > 1. \end{cases} \quad (14)$$

Compared with the original GARCH-MIDAS mode, the Model (14) clearly reveals how low-frequency variables directly affect the conditional variance of the high-frequency dependent variable. Specially, the impacts of low-frequency variables on the volatility on the first day in a month have a different pattern from their impacts on the other trading days. We find that the impacts of the low-frequency variables are reflected by the monthly adjusted intercepts of  $\gamma \exp\left(\theta \sum_{k=1}^K \varphi_k(\omega) Z_{t-k}^{l(v)}\right)$ . In addition, the slopes remain unchanged as  $\alpha$  and  $\beta$  since the second trading day in a month. However, for the first day in a month, such impacts can also change the slopes as  $\alpha$  and  $\beta$  multiplied by  $\exp\left(\theta \sum_{k=1}^K \varphi_k(\omega) \Delta Z_{t-k}^{l(v)}\right)$  respectively.

### 2.3. The estimation of VaR

To analyze the response of market risk or VaR to low-frequency information, we can easily obtain the quantile expression of Eq. (1) as follows:

$$Q_{r_{i,t}}(\xi | \Omega_{t-1}) = \mu + \sigma_{it} Q_{\varepsilon_{i,t}}(\xi), \quad (15)$$

where  $Q_{r_{i,t}}(\xi | \Omega_{t-1})$  is the quantile or VaR of return  $r_{it}$  at a  $\xi$  level.  $Q_{\varepsilon_{i,t}}(\xi)$  is the quantile of a standardized skewed t distribution at a  $\xi$  level, which remains unchanged with a fixed  $\xi$ .

Thus, the dynamic market risk is linearly correlated with the time-varying volatility  $\sigma_{it}$ . We put forth Eqs. (13) and (14) to explain how low-frequency variables impact the volatility and VaR respectively. To verify the accuracy of our risk measures, we use several different backtesting methods discussed in the empirical section.

## 3. Risky factors and data description

Soy-related products play an important role in the lives of Chinese people. China is the world's largest soybean consumer market, but has a low self-sufficiency rate and relies heavily on foreign imports. 100,328,200 tons of soybeans were imported in cumulative months throughout 2020, an increase of 11.7% year-over-year, with imports reaching \$ 39.528 billion, thus China is keen to manage the soybean market risk well and guarantee the supply chain security of soybean-related industries. In the empirical study, we focus on the the representative Soybean No. 1 Futures in Dalian Commodity Exchange. The Dalian Commodity Exchange is the largest agricultural futures market in China and



the second-largest soybean futures market in the world. Chinese Soybean No. 1 Futures uses the food-quality non-GMO soybeans as the underlying asset, and imported non-GMO soybeans of acceptable quality can be used as standard or substitutes for delivery. The logarithmic returns calculated from the closing price of Soybean No. 1 Futures (SFP) is adopted as the dependent variable. We next turn to systematically identify the risk factors from a comprehensive economic perspective.

### 3.1. Risky factors

Generally speaking, agricultural futures price fluctuations are mainly affected by a combination of factors such as supply and demand, inventories, substitutes, freight rates, trading positions and volumes, spot prices, price fluctuations in international markets, and macroeconomic trends.

*First*, supply and demand are the most direct market drivers. Soybean stocks act as a buffer and likewise act as a moderator for prices. The underlying factors behind commodity prices has been found from different perspectives such as both supply-demand and macroeconomic fundamentals (Akram, 2009; Matesanz, Torgler, Dabat, & Ortega, 2014), uncertainty, geopolitical risk, and extreme events (Antonakakis, Gupta, Kollias, & Papadamou, 2017; Balcilar, Gupta, & Pierdzioch, 2016; Bilgin, Gozgor, Lau, & Sheng, 2018; Gkillas, Gupta, & Pierdzioch, 2020), and speculative activity (Bohl, Siklos, & Wellenreuther, 2018). Among the early studies, the quality and quantity of supply are proved to have a great influence on commodity pricing (Chambers & Bailey, 1996; Deaton & Laroque, 1996; Mackey, 1989). Additionally, the rapid growth of commodity demand drives the rise in commodity prices (Cevik & Sedik, 2014; Kilian, 2009), even has a stronger influence than supply factors (Jacks & Stuermer, 2020). From the perspective of supply, demand and stock, we select five variables: China soybean import volume (IMPORT), China soybean output volume (OUTPUT), China soybean consumption (CONSUME), the amount for oil extraction (OILE), and the China soybean stock (STOCK).

*Second*, the information spillover among commodities is a core driver of commodity prices (Matesanz et al., 2014), so changes in the prices of downstream soybean products namely China soybean meal futures price (MEAL) and China soybean oil futures price (SOFP) can trigger changes in demand and have a knock-on effect on soybean futures volatility. The connections among commodities may source from the mutual substitution between agricultural products and fossil energy (Baffes, 2007), commodity financialization (Tang & Xiong, 2012), and common macroeconomic factors (Gleich, Achzet, Mayer, & Rathgeber, 2013). Thus, changes in the prices of China corn futures price (CORN) as a major soybean substitute will undoubtedly have a more direct impact on them.

*Third*, Natanelov, Alam, McKenzie, and Van-Huylenbroeck (2011) indicate that agricultural commodity futures price exhibits a long-term co-movement with crude oil prices. Liu, Pan, Yuan, and Chen (2019) find that 11 kinds of agricultural commodity futures prices present positive correlations with crude oil futures prices. transportation costs have a significant impact on commodity futures prices, and shipping costs are directly impacted by international crude oil prices. We use the Brent crude oil spot price (BRENT) as a risk factor. In addition, we also use the Baltic Dry Index (BDI\*), which reflects spot freight rates on major routes.

*Fourth*, due to the strong demand in the domestic soybean market, price fluctuations in the Chicago Board of Trade, which serve as a market vane, can be transmitted to the Chinese soybean market. In the short term, domestic spot prices as well as trading volume and trading positions in the futures market will have a more direct impact on futures prices. Therefore, we select four short-term risky factors namely the soybean futures closing price in the Chicago Board of Trade (CBOT), China soybean spot price (SPOT), the trading volume of China soybean (VOLUME), and positions of China soybean (POSITION).

*Fifth*, in the long term, the macroeconomic situation affects commodity futures price trends. Batten, Ciner, and Lucey (2010) suggest the spillover effects of macroeconomic determinants on commodity prices, including the business cycle, monetary environment, and financial market sentiment. The exchange rate of the US dollar as well as interest rates (Chiang, Chen, & Huang, 2019; Gruber & Vigfusson, 2018), economic activity (Klotz, Lin, & Hsu, 2014), and stock levels and yields (Balcombe, 2011) are also important sources to drive commodity price fluctuations. Recently, using a generalized dynamic factor model, Kagraoka (2016) suggests the US inflation rate, world industrial production, the world stock index, and the price of crude oil are four common dynamic factors that determine the commodity price. Uncertainty and extreme events are also a source of volatility. Joets, Mignon, and Razafindrabe (2017) find the agricultural and industrial markets are more sensitive to changes in macroeconomic uncertainty. Hamadi, Bassil, and Nehme (2017) propose that macroeconomic news surprise has significant impacts on the volatility of agricultural commodities. Prokopczuk, Stancu, and Symeonidis (2019) discover credit risk, financial market stress, and fluctuations in business conditions are important predictors of commodity volatility. Some researchers such as (Tzeng & Shieh, 2016) and (Vercammen, 2020) examine the performance of commodity markets in extreme cases and find clear differences of commodity markets in normal periods. Furthermore, Hu, Zhang, Ji, and Wei (2020) show how macro factors contribute to the volatility of soybeans, gold, and West Texas Intermediate (WTI) markets through a dynamic connectedness network. Based on these analyses, we choose four macro variables as risk factors: the exchange rate (ER), China consumer price index (CPI), China agricultural production price index (APPI) and China money supply (M2).

### 3.2. The data descriptive analysis

The data set is obtained from China's WIND financial database and ranges from 16 May 2008 to 24 December 2019. We consider 18 risky factors as discussed above. Table 1 reports their type, frequency, and definition. For daily variables, we take the form of logarithmic return or increment. In particular, the sample data is divided into two subgroups. The in-sample data is used for volatility modeling, covering from May 2008 to December 2017. The out-of-sample data from January 2018 to December 2019 remain for prediction.

Table 2 reports the descriptive statistics. We have 2652 daily observations and 140 monthly observations. The negative skewness for SFP indicates that its empirical distribution is left-skewed. We find the kurtosis of SFP is larger than 3 and all Jarque-Bera statistics are significant, which indicates the empirical distribution has a sharp peak and thick tails than the normal distribution. It supports us to use the skewed t distribution instead of normal distribution.

**Table 1.** Variable definition.

Variables	Types	Frequency	Definitions
SFP	Future price	Daily	The closing price of the Soybean No. 1 Future in Dalian Commodity Exchange
CBOT	Future price	Daily	The soybean futures closing price in the Chicago Board of Trade
SPOT	Spot Price	Daily	China soybean spot price
MEAL	Downstream product price	Daily	Soybean meal futures closing price in Dalian Commodity Exchange
SOFP	Downstream product price	Daily	Soybean oil futures closing price in Dalian Commodity Exchange
CORN	Substitute price	Daily	Corn futures closing price in Dalian Commodity Exchange
BDI	Freight rate	Daily	The Baltic dry index
ER	Exchange rate	Daily	The USD/CNY exchange rate
VOLUME	Market trading volume	Daily	Trading volume of Soybean No. 1 futures in Dalian Commodity Exchange
POSITION	Market positions	Daily	Positions of Soybean No. 1 futures in Dalian Commodity Exchange
IMPORT	Supply	Monthly	China soybean import volume
OUTPUT	Supply	Monthly	China soybean output
CONSUME	Demand	Monthly	China soybean consumption
OILE	Demand	Monthly	The amount of soybeans used for oil extraction
CPI	Consumer price index	Monthly	China consumer price index on a month-over-month basis
APPI	The agricultural production price index	Monthly	China agricultural production price index on a month-over-month basis
M2	Money supply	Monthly	China money supply
BRENT	Crude oil market	Monthly	Spot fob price of Brent crude oil
STOCK	Stock	Monthly	China soybean stocks at the end of a month

**Table 2.** Descriptive statistics of the variables.

Variables	Obs.	Mean	Max.	Min.	Std. Dev.	Skewness	Kurtosis	Jarque-Bera (P-value)
SFP	2652	−0.0002	0.1128	−0.0997	0.0128	−0.1422	13.3314	<b>11,803.4263</b> (0.00%)
CBOT	2652	−0.0001	0.2032	−0.2819	0.0180	−2.4840	46.3707	<b>210,579.6751</b> (0.00%)
SPOT	2652	−0.0001	0.0679	−0.0996	0.0056	−2.3602	83.1174	<b>711,739.1253</b> (0.00%)
MEAL	2652	−0.0001	0.1305	−0.2089	0.0179	−1.1832	19.6249	<b>31,159.6296</b> (0.00%)
SOFP	2652	−0.0002	0.2321	−0.1920	0.0186	1.3392	32.7266	<b>98,438.6026</b> (0.00%)
CORN	2652	3.45E-05	0.1419	−0.1909	0.0133	−0.7336	43.5635	<b>182,054.2615</b> (0.00%)
BDI	2652	−0.0009	0.1939	−0.2247	0.0284	0.0012	9.1868	<b>4229.4957</b> (0.00%)
ER	2652	−1.52E-06	0.0184	−0.0113	0.0016	0.6863	19.6133	<b>30,706.1925</b> (0.00%)
VOLUME	2652	−0.0012	3.9722	−4.9824	0.4456	0.2007	12.2264	<b>9424.2302</b> (0.00%)
POSITION	2652	−0.0008	0.4578	−0.8324	0.0523	−0.7231	36.4095	<b>123,570.5922</b> (0.00%)
IMPORT	140	5.7632	10.0800	2.1300	1.9133	0.3539	2.1818	<b>6.8283</b> (0.00%)
OUTPUT	140	13.9131	16.3500	11.7900	1.3829	−0.2200	1.5989	<b>12.5805</b> (0.00%)
CONSUME	140	73.4781	106.3000	44.5400	19.4456	0.1648	1.8187	<b>8.7739</b> (0.00%)
OILE	140	61.5946	90.0000	34.5000	17.7372	0.0717	1.7573	<b>9.1279</b> (0.00%)
CPI	140	2.4453	7.7000	−1.8000	1.7047	0.3412	4.1932	<b>11.0203</b> (0.00%)
APPI	140	3.5076	24.8000	−7.5100	6.0632	1.6128	6.0975	<b>116.6616</b> (0.00%)
M2	140	14.1058	29.7400	8.0000	5.3113	1.2764	4.3166	<b>48.1266</b> (0.00%)
BRENT	140	79.0717	133.9000	30.8000	26.5656	0.2178	1.7275	<b>10.5527</b> (0.00%)
STOCK	140	13.3867	23.5200	2.6600	5.6366	−0.4477	2.5350	5.9383 (0.00%)

The bold estimators are statistically significant at the 5% level.

The correlations between variables are reported in [Table 3](#). Two factors are strongly correlated if the correlation coefficient exceeds 0.5 and is statistically significant at the 5% level. We find that 39% of all the coefficients show a strong correlation. To avoid multicollinearity, only the factor with the largest variance contribution from a set of strongly correlated factors is retained.

**Table 3.** Correlations between variables.

Daily variables	CBOT	SPOT	MEAL	SOFP	CORN	BDI	ER	VOLUME	POSITION
CBOT	1.0000								
	—								
SPOT	<b>0.8720</b>	1.0000							
	0.00%	—							
MEAL	<b>0.7551</b>	<b>0.8536</b>	1.0000						
	0.00%	0.00%	—						
SOFP	<b>0.7446</b>	<b>0.6570</b>	<b>0.5349</b>	1.0000					
	0.00%	0.00%	0.00%	—					
CORN	<b>0.5866</b>	<b>0.5147</b>	<b>0.3019</b>	<b>0.3016</b>	1.0000				
	0.00%	0.00%	0.00%	0.00%	—				
BDI	<b>0.1484</b>	<b>0.3347</b>	<b>0.3810</b>	<b>0.4306</b>	−0.2712	1.0000			
	0.00%	0.00%	0.00%	0.00%	0.00%	—			
ER	−0.4863	−0.4248	−0.2088	−0.0556	−0.7983	<b>0.3825</b>	1.0000		
	0.00%	0.00%	0.00%	0.42%	0.00%	0.00%	—		
VOLUME	<b>0.0911</b>	<b>0.1917</b>	<b>0.2524</b>	<b>0.2339</b>	−0.2112	<b>0.3437</b>	<b>0.2238</b>	1.0000	
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	—	
POSITION	<b>0.4702</b>	<b>0.5020</b>	<b>0.3883</b>	<b>0.6056</b>	<b>0.3001</b>	<b>0.3172</b>	−0.1418	<b>0.4532</b>	1.0000
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	—
Monthly variables	IMPORT	OUTPUT	CONSUME	OILE	CPI	APPI	M2	BRENT	STOCK
IMPORT	1.0000								
	—								
OUTPUT	−0.5752	1.0000							
	0.00%	—							
CONSUME	<b>0.7712</b>	−0.5983	1.0000						
	0.00%	0.00%	—						
OILE	<b>0.7758</b>	−0.6173	<b>0.9993</b>	1.0000					
	0.00%	0.00%	0.00%	—					
CPI	−0.1266	<b>0.3991</b>	−0.1553	−0.1612	1.0000				
	13.62%	0.00%	6.70%	5.71%	—				
APPI	−0.2750	<b>0.5102</b>	−0.2990	−0.3111	<b>0.7668</b>	1.0000			
	0.10%	0.00%	0.03%	0.02%	0.00%	—			
M2	−0.6442	<b>0.4110</b>	−0.8175	−0.8225	−0.2142	−0.1200	1.0000		
	0.00%	0.00%	0.00%	0.00%	1.11%	15.78%	—		
BRENT	−0.3981	<b>0.5655</b>	−0.4252	−0.4224	<b>0.5735</b>	<b>0.4301</b>	0.1340	1.0000	
	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.44%	—	
STOCK	<b>0.7001</b>	−0.3803	<b>0.9112</b>	<b>0.9126</b>	0.0398	−0.2146	−0.8369	−0.1533	1.0000
	0.00%	0.00%	0.00%	0.00%	64.09%	1.09%	0.00%	7.06%	—

P values are below the correlation. Bold estimators are statistically significant at the 5% level.

## 4. Empirical results

In this section, we demonstrate in detail the process of measuring China's soybean futures market risk using the new model, presenting the key empirical findings.

### 4.1. Identifying important factors

We perform a data pre-processing for each risk factor. The level and variance of monthly factors are calculated using the error and squared error respectively from an auto-regressive process (Engle et al., 2013). For daily factors, we use its monthly realized volatility to explain the long-term component of soybean volatility. Engle et al. (2013) adopt the variance ratios, that is,  $VaR(\log(\tau_t))/VaR(\log(\tau_{t,i}))$ , to measure the contributions of economic sources to the expected volatility. We can use it to evaluate the importance of factors for explaining volatility. Table 4 reports the variance ratios of the level and variance of all factors.

The variance ratios obtained by level values of factors are higher than those using variances of the factors, except corn futures price, money supply, consumer price index, and market trading volume. Each factor has two variance ratios, and we define the maximum one as its contribution. We rank all factors in the descending order of variance ratio, that is, Brent(level), CORN(variance), SOFP(level), MEAL(level), ER(level), POSITION(level), OUTPUT(level), CBOT(level), SPOT(level), CPI(variance), M2(variance), IMPORT(level), VOLUME(variance), CONSUME(level), OILE(level) and BDI(level). The last two factors (APPI and STOCK) have weaker contributions with variance ratios lower than 2%.

Further, Table 5 reports the estimators in all single-factor GARCH-MIDAS-Skewed  $t$  models. The estimators of  $\theta$  except four factors namely OILE, BDI, APPI, and STOCK are statistically significant at a 5% level. For the related international commodity markets, we find that: (1) the Brent crude oil price (BRENT) produces a negative impact on the volatility of SFP because it directly affects the production and transportation costs of soybeans; (2) the international soybean futures price (CBOT), as a bellwether, has a positive impact on the volatility of SFP; (3) the position (POSITION) and market trading volume (VOLUME) have significant negative impacts on it, but the Chinese spot soybean price level (SPOT) plays an opposite role. For interested substitutes and downstream products, the rising prices of corn futures (CORN), soybean oil (SOFP), and soybean meal futures (MEAL) will significantly increase the volatility of SFP.

On the side of supply and demand, it is found that: (1) the higher domestic soybean yields (OUTPUT) in China, the less volatility of SFP. However, the more volatility of soybean import volume (IMPORT), the more volatility of SFP. This finding is questionable because China's soybean imports are mainly GMO soybeans, but also contain a small proportion of non-GMO soybeans and the proportion is increasing in recent years, while the underlying of the SFP is non-GMO soybeans, so we speculate that as the volume of imported soybeans expands, especially non-GMO soybeans, it will help alleviate demand for edible-grade non-GMO soybeans and thereby reduce price volatility. Therefore, we

**Table 4.** The variance ratios of all single-factor models.

Types	Variables	Level value	Variance value
Crude oil market	BRENT	<b>50.29%</b>	9.72%
Substitute price	CORN	1.93%	<b>38.29%</b>
Downstream product price	SOFP	<b>34.82%</b>	22.08%
Downstream product price	MEAL	<b>30.93%</b>	14.41%
Exchange rate	ER	<b>24.42%</b>	0.02%
Market positions	POSITION	<b>23.84%</b>	10.69%
Supply	OUTPUT	<b>22.92%</b>	0.03%
Future price (CBOT)	CBOT	<b>20.53%</b>	6.73%
Spot Price	SPOT	<b>18.38%</b>	0.90%
Consumer price index	CPI	3.39%	<b>15.56%</b>
Money supply	M2	0.26%	<b>14.43%</b>
Supply	IMPORT	<b>12.19%</b>	0.74%
Market trading volume	VOLUME	8.68%	<b>12.09%</b>
Demand	CONSUME	<b>5.58%</b>	0.06%
Demand	OILE	<b>5.30%</b>	0.17%
Freight rate	BDI	<b>4.88%</b>	3.40%
Agricultural production price index	APPI	<b>1.90%</b>	0.59%
Chinese soybean stock	STOCK	<b>0.44%</b>	0.09%

Bold letters indicate the maximum of variance ratios among level and variance effects.

**Table 5.** Estimators of all single-factor GARCH-MIDAS-Skewed t models.

Variables	Effects	$\mu$	$\alpha$	$\beta$	$m$	$\theta$	$\omega$	$g_0$	$\eta$	$\lambda$	BIC	NLLF	K
BRENT	Level	-0.0001	0.9996	8.24E-13	0.8109	-0.0199	1.0000	3.2541	2.1798	-0.0020	-6.6267	-6770.8109	9
	Variance	57.22%	0.00%	100.00%	77.45%	<b>166.4403</b>	8.73%	4.12%	6.07%	3.46%			
CORN		-0.0001	0.9996	0.0000	-1.2761		1.0000	0.2511	2.1830	-0.0118	-6.6154	-6759.3663	9
	Variance	19.48%	0.00%	100.00%	32.72%	0.00%	2.07%	5.67%	8.51%	4.09%			
SOFP	Level	-0.0001	0.9986	0.0000	-1.9605	<b>19.5072</b>	1.8385	1.1927	2.1915	-0.0096	-6.6167	-6760.6273	9
	Variance	22.86%	0.00%	100.00%	18.99%	0.00%	4.34%	2.13%	9.05%	6.76%			
MEAL	Level	-0.0002	0.9991	0.0000	-1.6303	<b>12.6872</b>	2.9767	1.5801	2.1937	-0.0185	-6.6130	-6756.8765	9
	Variance	6.75%	0.00%	100.00%	13.00%	0.00%	6.71%	4.69%	7.30%	5.03%			
ER	Level	-0.0001	0.9990	-3.52E-18	-1.5682	<b>91.1542</b>	1.0000	5.4274	2.1859	-0.0098	-6.6071	-6750.8523	9
	Variance	24.05%	0.00%	100.00%	34.65%	0.00%	9.43%	2.58%	6.14%	4.72%			
POSITION	Level	-0.0001	0.9968	0.0000	-2.8132	-5.1123	1.0000	3.5621	2.1816	-0.0100	-6.6068	-6750.6115	9
	Variance	19.10%	0.00%	100.00%	4.56%	0.00%	6.68%	1.29%	4.81%	9.53%			
OUTPUT	Level	-0.0001	0.9993	-2.62E-06	1.6919	-0.2179	5.7410	1.6159	2.1853	-0.0062	-6.6065	-6750.3001	9
	Variance	37.31%	0.00%	99.94%	51.62%	0.00%	8.04%	3.87%	6.90%	1.56%			
CBOT	Level	-0.0002	0.9991	0.0000	-1.5628	<b>11.5692</b>	1.1258	1.1631	2.1945	-0.0190	-6.6060	-6749.8044	9
	Variance	6.78%	0.00%	100.00%	0.00%	0.00%	1.98%	6.34%	4.47%	3.03%			
SPOT	Level	-0.0002	0.9994	0.0000	-1.2647	<b>16.9994</b>	5.5727	0.2060	2.1990	-0.0137	-6.6054	-6749.2119	9
	Variance	12.90%	0.00%	100.00%	43.40%	0.00%	6.19%	4.63%	5.90%	2.36%			
CPI	Variance	-0.0001	0.9996	0.0000	-0.6164	-0.6996	55.1109	0.8415	2.1856	-0.0104	-6.6030	-6746.7056	9
	Variance	20.91%	0.00%	100.00%	35.90%	0.05%	3.46%	5.95%	2.05%	4.88%			
M2	Variance	-0.0002	0.9957	0.0000	-3.2961	<b>0.1805</b>	8.8088	0.2037	2.1931	-0.0124	-6.6023	-6746.0537	9
	Variance	16.10%	0.00%	100.00%	84.11%	0.04%	9.72%	8.25%	3.68%	2.96%			
IMPORT	Level	-0.0001	0.9997	5.24E-13	-1.3522	<b>0.1344</b>	6.4223	0.0060	2.1918	-0.0069	-6.6018	-6745.4968	9
	Variance	33.07%	0.00%	100.00%	48.78%	1.29%	5.22%	9.16%	4.08%	6.23%			
VOLUME	Variance	-0.0001	0.9993	0.0000	-0.7586	-0.1439	1.1935	1.3937	2.1945	-0.0092	-6.6013	-6745.0253	9
	Variance	22.97%	0.00%	100.00%	21.68%	0.27%	1.68%	6.47%	8.46%	3.12%			
CONSUME	Level	-0.0001	0.9995	4.46E-12	-1.5265	<b>0.0093</b>	55.4894	0.1102	2.1895	-0.0100	-6.5985	-6742.1338	9
	Level	21.28%	0.00%	100.00%	27.87%	3.46%	2.92%	5.16%	9.29%	5.53%			
OILE	Level	-0.0001	0.9996	0.0000	-1.1602	0.0097	51.3803	0.0090	2.1895	-0.0101	-6.5984	-6742.0304	9
	Level	21.16%	0.00%	100.00%	73.50%	17.93%	3.59%	8.44%	6.73%	1.15%			
BDI	Level	-0.0002	0.9996	0.0000	-0.8499	-0.4944	54.0268	0.4720	2.1909	-0.0119	-6.5983	-6741.9388	9
	Level	16.82%	0.00%	100.00%	67.09%	9.61%	4.51%	9.87%	2.23%	3.06%			
APPI	Level	-0.0002	0.9996	-4.91E-07	-0.6092	-0.0170	34.0419	0.5180	2.1901	-0.0115	-6.5971	-6740.7352	9
	Level	17.00%	0.00%	99.98%	72.16%	20.32%	9.66%	6.23%	4.04%	7.95%			
STOCK	Level	-0.0002	0.9994	0.0000	-1.2974	0.0087	45.5104	0.2179	2.1924	-0.0114	-6.5965	-6740.0902	9
	Level	17.23%	0.00%	100.00%	63.72%	58.34%	8.35%	7.03%	9.14%	6.26%			

P values are below the estimators. Bold estimators are statistically significant at the 5% level.

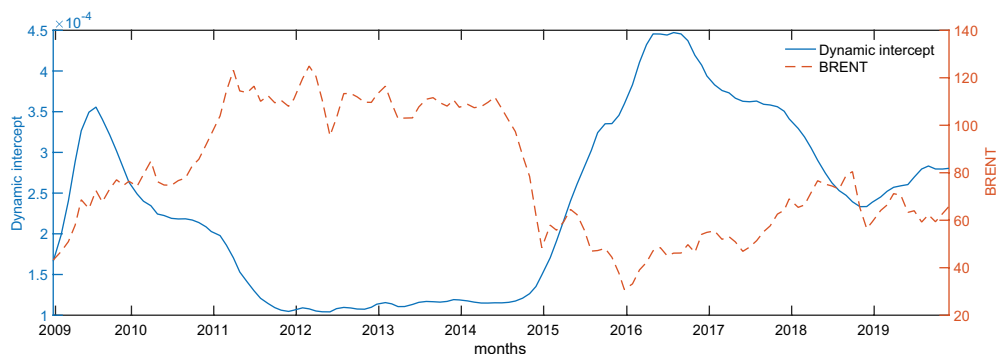
will re-examine this relationship in the multi-factor model by introducing additional control variables. (2) the impacts of soybean consumption (CONSUME) and demand for crushing oil (OILE) are both positive but much weaker than those of soybean supply.

In terms of macro factors, it is shown that: (1) the variance of the consumer price index (CPI) has a significant negative impact on the volatility of SFP, but the variance of money supply (M2) contributes positively to it; (2) the level of freight rate(BDI) has a negative impact with a low variance ratio of 4.88%. The APPI and STOCK are much weaker and can be ignored.

#### 4.2. The direct impacts of low-frequency factors on the market risk

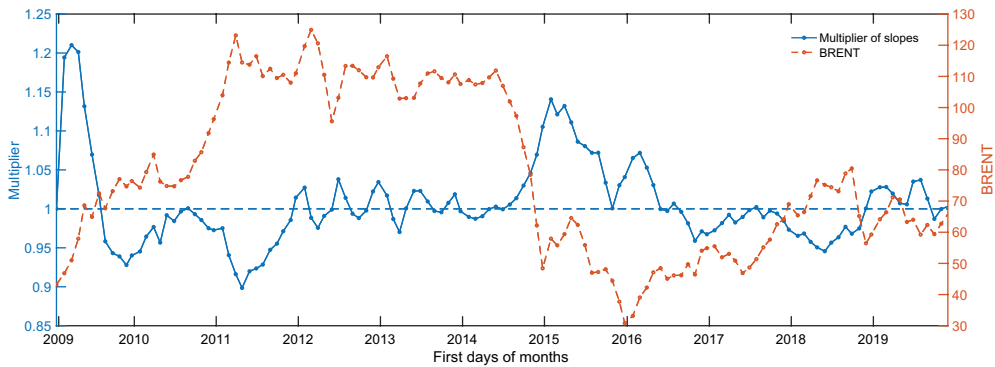
Based on the Model (14) and Eq. (15), we discuss the dynamic path on low-frequency factors influencing the high-frequency volatility or market risk. To be concise, we take the most influential factor of BRENT level as an example to analyze the time-varying impact mechanism. We plot the dynamic intercepts in each month and dynamic multipliers of slopes only on the first day of each month in Figures 1 and 2 respectively.<sup>1</sup> According to the results in Table 5, we find that the estimated coefficient of  $\theta$  is negative for BRENT level factor ( $Z^l$ ) and that  $\gamma = (1 - \alpha - \beta)\exp(m)$  and the weights  $\varphi(\omega)$  are always positive, so the dynamic intercept  $\gamma\exp(\theta\sum_{k=1}^K\varphi_k(\omega)Z_{t-k}^l)$  is nonlinearly negatively correlated with the BRENT level factor as showed in Figure 1. For instance, the intercept is lower during the high oil price period of 2011–2014 and higher during the low and sluggish oil price period of 2016–2018.

The multiplier of slope of  $\exp(\theta\sum_{k=1}^K\varphi_k(\omega)\Delta Z_{t-k}^l)$  on the first day of each month in Eq. (14) also has a non-linear negative correlation with the MIDAS-weighted increments  $\sum_{k=1}^K\varphi_k(\omega)\Delta Z_{t-k}^l$ . We are more interested in whether the multiplier of slope is greater than one. In fact, since the weight  $\varphi(\omega)$  is positive, it depends mainly on the crude oil price change series  $\Delta Z_t^l$ . The slope multiplier is less than one when the price of crude oil continues to rise and greater than one when it continues to fall. This can be confirmed from Figure 2. The multiplier of slope is less than one during the sustained high oil price climbing phase in 2010–2012, while it is greater than one during the sustained declining oil price phase in 2014–2016.



**Figure 1.** The dynamic relationship between the time-varying intercepts  $\gamma\exp(\theta\sum_{k=1}^K\varphi_k(\omega)Z_{t-k}^l)$  and the BRENT crude oil price. The latter is plotted in the right vertical axis.





**Figure 2.** The dynamic relationship between the multipliers of slope on the first day of each month  $\exp(\theta \sum_{k=1}^K \varphi_k(\omega) \Delta Z_{t-k}^I)$  and BRENT crude oil price. The latter is plotted in the right vertical axis.

In short, in the long run, high Brent oil prices have a lower impact on the volatility of Chinese soybean futures returns or market risk than low oil prices, perhaps depending on the asymmetric response of investors in the Chinese soybean futures market to expectations about the correlation between international oil price movements and Chinese soybean futures returns, that is, low oil prices could reduce the transportation costs and prices of imported soybeans, which in turn could put pressure on Chinese soybean futures prices and trigger its strong volatility, and vice versa. We can perform similar analysis to other factors, which is omitted here.

#### 4.3. The multi-factor GARCH-MIDAS-Skewed $t$ model

Because multiple factors of different frequencies simultaneously impact soybean futures volatility, it is necessary to construct a GARCH-MIDAS-Skewed  $t$  model with more than one factor. As showed in Table 3, the correlations between factors are pretty high, which might cause a multicollinearity issue. We take two simple steps to overcome this problem. First, if several factors are highly correlated, we will only keep the factor with the highest variance ratio. For instance, SPOT, MEAL, and CBOT are highly correlated with SOFP (over 0.7), and the correlation between CONSUME and OILE reaches 0.9993. Therefore, we can exclude SPOT, MEAL, CBOT, and OILE in the multi-factor GARCH-MIDAS-Skewed  $t$  model, because their variance ratios are relatively lower. Second, the stepwise regression is used to select variables by adding one factor at a time in the order of variance ratios from the highest to the lowest. Specifically, a new factor only can be retained when it is statistically significant and raises the variance ratio. Finally, ten factors are remained namely BRENT(level), CORN(variance), SOFP(level), MEAL(level), CPI(variance), M2(variance), IMPORT(level), VOLUME(variance), CONSUME(level) and BDI(level). The estimated results of the multi-factor GARCH-MIDAS-Skewed  $t$  model are shown in Table 6.

The estimated dynamic long-term component of volatility and daily conditional variances of soybean futures price are shown in Figure 3. There is a strong consistency in the trend between estimated long-term volatility and daily volatility in soybean futures. Overall, the multi-factor GARCH-MIDAS-Skewed  $t$  model has a higher



variance ratio of 80.79% (seen in Table 6) than the maximum 50.29% among all single-factor models (seen in Table 4), which provide stronger contributions to the volatility of SFP. The results indicate that international crude oil prices, prices of downstream products, consumer price index, money supply, China's soybean imports, the volume of China's soybean futures trading, China soybean consumption, and freight rate are significant determinants of the long-term volatility of China soybean futures.

#### 4.4. Backtesting of VaR

As a new parametric model in measuring market risk, the multi-factor GARCH-MIDAS-Skewed t model works well to analyze the volatility of Chinese soybean futures. We next evaluate its performance in predicting VaR, compared with the GARCH and GARCH-MIDAS (Engle et al., 2013) models with a normal distribution. We focus on assessing the out-of-sample predictive power, covering 458 daily samples from January 2018 to December 2019.

Market positions in the soybean futures market are divided into short and long positions. Holding a long position will result in losses when the futures price falls, while holding a short position will result in losses when the price rises. Our model uses a skewed t distribution that can well describe the asymmetry in the distribution of soybean futures returns and therefore allows for a precise assessment of market risk for both long and short position assets. When an asset return at the long (short) position is lower (higher) than the estimated VaR, it happens a VaR violation or a hit. The hit ratio (HR) is a simple measure of predictive quality, which is defined as the number of hits divided by the total number of samples. A good model should have an HR close to the significance level (P) of VaR, thus we define a relative percentage index as  $PI = \frac{|HR-P|}{P}$ . The less PI, the better the model.

The backtesting of VaR forecasts is usually performed by likelihood ratio (LR) tests proposed by (Kupiec, 1995) and (Christoffersen, 1998), which assesses the unconditional coverage (UC) and conditional coverage (CC) as well as the independence (ind) of VaR exceedances or violations. Then, we also adopted the popular dynamic quantile (DQ) test proposed by (Engle & Manganelli, 2004), which links the violations to a set of explanatory variables that include a constant, the VaR forecast and the first four lagged hits.

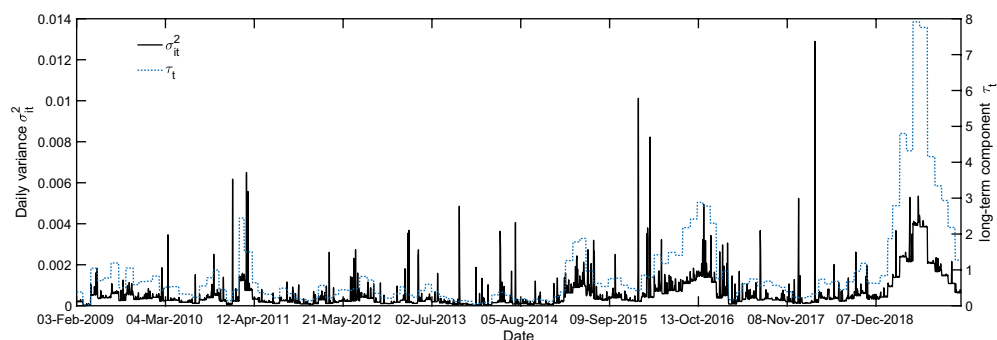
We compare the forecasting performances of the interested models, considering long and short positions and two significance levels (1% and 5%). The multi-factor GARCH-MIDAS-normal and multi-factor GARCH-MIDAS-Skewed t models have the same explanatory variables in Section 4.3. The backtesting of out-of-sample VaR forecasts and the model comparison are shown in Table 7.

First, each model has four PIs in Table 7. We calculate their means that are 53.57%, 45.27%, and 34.80% for GARCH(1,1), GARCH-MIDAS-Normal, and GARCH-MIDAS-Skewed t models, respectively. The GARCH and GARCH-MIDAS models with a normal distribution do not yield accurate forecasts, especially at the 5% level. Therefore, our model with the least PI value has a better performance in VaR forecasts than them. In particular, our model performs better for short VaR forecasts than for long VaR forecasts, because the PIs are lower in the right tail than in the left tail.

**Table 6.** Estimators of the multi-factor GARCH-MIDAS-Skewed t model.

Variables	$\mu$	$\alpha$	$\beta$	$m$	$\theta_1$ (BRENT)	$\theta_2$ (CORN)	$\theta_3$ (SOFP)	$\theta_4$ (MEAL)	$\theta_5$ (CPI)	$\theta_6$ (M2)
Coefficients	-0.0001	0.9993	-1.43E-06	2.2808	-0.0254	18.6396	12.3816	8.7480	-1.0666	0.5439
P-value	32.33%	0.00%	76.90%	10.49%	0.00%	6.79%	0.64%	0.04%	0.03%	0.00%
Variables	$\theta_7$ (IMPORT)	$\theta_8$ (VOLUME)	$\theta_9$ (CONSUME)	$\theta_{10}$ (BDI)	$\omega_1$	$\omega_2$	$\omega_3$	$\omega_4$	$\omega_5$	$\omega_6$
Coefficients	-1.0923	-0.3801	0.0846	-4.3560	310.1599	1.0000	1.7543	2.8076	1.8013	210.2896
P-value	0.00%	0.00%	0.00%	0.00%	2.84%	5.93%	2.65%	7.16%	5.47%	1.03%
Variables	$\omega_7$	$\omega_8$	$\omega_9$	$\omega_{10}$	$g_0$	$\eta$	$\lambda$	NULLF	variance ratio	
Coefficients	2.3862	9.4413	3.4755	1.0000	9.3603	2.1518	-0.0088	-6826.4282	80.79%	
P-value	9.27%	3.59%	4.39%	6.70%	3.95%	6.82%	5.12%			

Bold estimators are statistically significant at the 5% level.



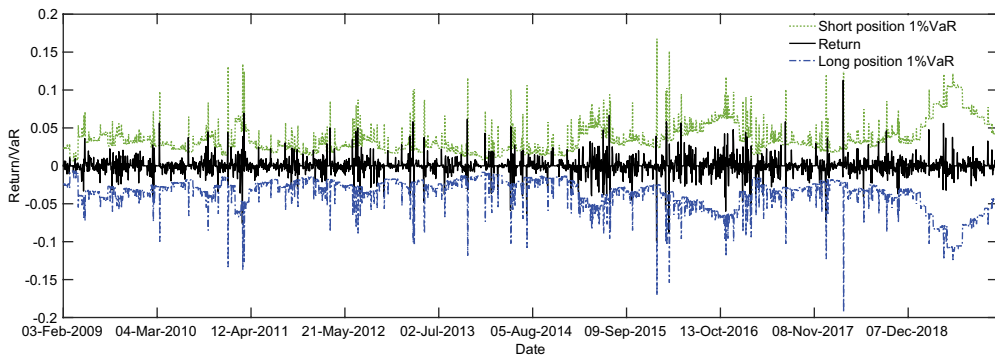
**Figure 3.** The estimated dynamic long-term component of volatility and daily conditional variances of the soybean futures price.

*Second*, it should be noted that to distinguish between long and short positions, our model allows VaR to be negative, although in practice it is customary to express risk in terms of the amount of VaR. In the financial industry, according to the Basel Accord, capital adequacy depends on the absolute value of the predicted VaR. With the same quality of backtesting, models with lower absolute VaRs will have a more competitive advantage because they will achieve the same risk prevention effect by only having to maintain a lower level of capital adequacy. To assess the economic value of risk models, we

**Table 7.** Backtesting of out-of-sample VaR forecasts and the model comparison.

Models	Hit ratio	PI	MAVaR	LRuc	LRind	LRcc	DQ
Panel A: Long position, P = 0.01							
GARCH-Normal	0.0197	97.00%	0.0308	0.0378 (15.41%)	0.1326 (28.42%)	0.1704 (8.17%)	<b>21.1049</b> (0.18%)
GARCH-MIDAS-Normal	0.0087	12.66%	0.0365	0.0775 (21.93%)	0.0882 (23.35%)	0.1657 (7.95%)	<b>27.8746</b> (0.01%)
GARCH-MIDAS-Skewed t	0.0022	78.17%	0.0531	4.1448 (4.18%)	0.0088 (7.46%)	4.1536 (12.53%)	2.9366 (81.68%)
Panel B: Long position, P = 0.05							
GARCH-Normal	0.0153	69.43%	0.0234	<b>15.7809</b> (0.01%)	0.2486 (38.20%)	<b>16.0295</b> (0.03%)	<b>17.3645</b> (0.80%)
GARCH-MIDAS-Normal	0.0197	60.60%	0.0260	<b>11.4290</b> (0.07%)	0.4014 (47.36%)	<b>11.8304</b> (0.27%)	<b>23.4456</b> (0.07%)
GARCH-MIDAS-Skewed t	0.0633	26.64%	0.0136	1.5831 (20.83%)	4.0639 (4.38%)	5.6470 (5.94%)	11.0046 (8.82%)
Panel C: Short position, P = 0.01							
GARCH-Normal	0.0109	9.00%	0.0308	3.3627 (6.67%)	0.4014 (47.36%)	3.7641 (15.23%)	6.2086 (40.02%)
GARCH-MIDAS-Normal	0.0066	34.00%	0.0533	0.6270 (42.85%)	0.0528 (18.18%)	0.6798 (28.82%)	2.5391 (86.41%)
GARCH-MIDAS-Skewed t	0.0087	13.00%	0.0518	0.0775 (21.93%)	0.0882 (23.35%)	0.1657 (7.95%)	1.6225 (95.09%)
Panel D: Short position, P = 0.05							
GARCH-Normal	0.0306	38.86%	0.0234	4.2026 (4.04%)	0.9472 (33.04%)	5.1498 (7.62%)	4.8025 (56.94%)
GARCH-MIDAS-Normal	0.0131	73.80%	0.0376	<b>18.3756</b> (0.00%)	0.1861 (33.38%)	<b>18.5616</b> (0.01%)	14.8413 (2.15%)
GARCH-MIDAS-Skewed t	0.0393	21.40%	0.0228	1.1874 (27.58%)	1.5568 (21.21%)	2.7442 (25.36%)	5.2793 (50.85%)

P values are below the estimators. Bold estimators indicate rejections from the LR tests and DQ test at a 1% significance level.



**Figure 4.** The estimated VaR for long and short soybean futures assets at a 1% confidence level.

calculate the mean of the absolute value of out-of-sample VaRs (MAVaR) in each model as showed in Table 7. In Panel B, C, and D of Table 7, the GARCH-MIDAS-Skewed t model presents lower PIs and lower MAVaRs than the GARCH-MIDAS-Normal model. In Panel A, although GARCH-MIDAS-Normal model has a lower MAVaR than the GARCH-MIDAS-Skewed t model, it does not pass the DQ test at a 1% significance level.

*Third*, it is important for backtesting process that only the GARCH-MIDAS-Skewed t model has passed all LR test and DQ test at 1% and 5% significance levels both for long and short positions. We plot the predicted VaRs for soybean futures long and short positions at the 1% significant level using the multi-factor GARCH-MIDAS-Skewed t model in Figure 4.

Overall, the multi-factor GARCH-MIDAS-Skewed t model has obvious advantages over other investigated models in predicting VaR.

## 5. Conclusions

China has always experienced a very high market demand for soybeans and increased dependence on foreign soybean imports in the past decade. It has triggered dramatic fluctuations in soybean futures prices. Therefore, risk management in the soybean futures market has become extremely important for stakeholders, which is prevalent in other agricultural futures markets. As a result, there is an urgent need for more sophisticated risk measurement tools on a global scale. Traditional risk measurement models are flawed in their inability to use mixed frequency information. The contribution of this paper includes two aspects. We first put forth an improved GARCH-MIDAS-Skewed t model to overcome this dilemma as described in Section 2, which has many flexible settings as described in Section 2 than the GARCH-MIDAS model (Engle et al., 2013). The multi-factor mixed-frequency GARCH-MIDAS-Skewed t model has a higher variance ratio than all single-factor models. It can also produce better VaR forecasts for long and short positions compared to the GARCH and GARCH-MIDAS-normal models.

In addition, we provide a practical and instructive analysis of volatility and market risk in the Chinese soybean futures market. We believe this approach is equally applicable to other agricultural futures. We systematically demonstrate how to identify the important

factors, find the best combination of factors, and perform a backtesting of VaR forecasts. Based on the findings of the empirical analysis, we give five useful recommendations for risk management in the Chinese soybean futures market as follows:

(1) Investors should pay particularly close attention to the volatility of international crude oil prices. Brent crude oil prices are linked to soybean production and transportation costs, and when crude oil prices fall, soybean futures prices fall in tandem, with increased demand triggering increased volatility. The Baltic Dry Index freight rates play a similar role, but have a much weaker impact than oil prices.

(2) It is important to monitor price fluctuations of substitutes such as corn and downstream products such as soybean oil, which can trigger a homogeneous linkage of risks in the soybean futures market.

(3) The impacts of CPI and M2 cannot be ignored in the long term. Increased CPI volatility and higher price instability will increase investor uncertainty about commodity price expectations, thereby reducing market trading sentiment and price volatility. When the money supply fluctuates more, it is important to guard against increased risk in the soybean futures market.

(4) Market participants must pay attention to changes in the volume and structure of China's soybean imports (GMO and non-GMO soybeans). China's import dependence on soybeans has been above 80%, with China importing soybeans mainly to meet protein feed demand, and in recent years China has been increasing its imports of non-GMO soybeans to meet edible vegetable oil demand. More than 80% of China's homegrown soybeans are processed into food. As a result, there is a certain degree of substitutability between imported soybeans and domestically produced soybeans in terms of usage, so the increase in the amount of imported soybeans will, to a certain extent, reduce the volatility and market risk in China's soybean futures market. Of course, the increased consumption of soybeans in China will directly exacerbate volatility and market risk in the soybean futures market.

(5) China should develop a reasonable subsidy policy to encourage farmers to plant soybeans, increase the domestic supply of soybeans, and reduce dependence on foreign soybean imports; at the same time, expand the channels of soybean importing countries, establish reasonable trade mechanisms such as mutually lowering tariffs and increasing quotas; expand the proportion of imports of non-GMO soybeans, and maintain stable domestic soybean spot prices and futures prices.

In summary, we propose a new systematic framework for measuring futures market risk that integrates mixed-frequency data information into volatility modeling and market risk measurement, which can effectively address the scenario where the influencing factors have different observation frequencies and provide a method for identifying optimal factor combinations. In particular, combined with the empirical findings, we propose effective recommendations for improving risk management in the Chinese soybean futures market.

## Disclosure statement

No potential conflict of interest was reported by the authors.

## Funding

This work was supported by the National Natural Science Foundation of China [No.71871215, No. 71871120].

## Notes on contributors

**Xinyu Wang** is a professor in School of Economics and Management, China University of Mining and Technology. He specializes in financial engineering and risk management.

**Lele Zhang** is a full-time doctoral student in School of Business and Management, Shanghai International Studies University.

**Qiuying Cheng** is a full-time doctoral student in School of Economics and Management, China University of Mining and Technology.

**Song Shi** is a full-time master's degree student in School of Economics and Management, China University of Mining and Technology.

**Huawei Niu** is a professor in School of Economics and Management, China University of Mining and Technology. He specializes in financial engineering, credit risk and credit derivatives pricing.

## References

- Akram, Q. F. (2009). Commodity prices, interest rates and the dollar. *Energy Economics*, 31(6), 838–851.
- Antonakakis, N., Gupta, R., Kollias, C., & Papadamou, S. (2017). Geopolitical risks and the oil-stock nexus over 1898–2016. *Finance Research Letters*, 23, 165–173.
- Asgharian, H., Hou, A. J., & Javed, F. (2013). The importance of the macroeconomic variables in forecasting stock return variance: A GARCH-MIDAS approach. *Journal of Forecasting*, 32(7), 600–612.
- Baffes, J. (2007). *Oil spills on other commodities*. The World Bank.
- Balcilar, M., Gupta, R., & Pierdzioch, C. (2016). Does uncertainty move the gold price? New evidence from a nonparametric causality-in-quantiles test. *Resources Policy*, 49, 74–80.
- Balcombe, K. (2011). The nature and determinants of volatility in agricultural prices: An empirical study. In: A. Prakash (Ed.), *Safeguarding Food Security in Volatile Global Markets*, (pp. 89–110). Rome: FAO.
- Batten, J. A., Ciner, C., & Lucey, B. M. (2010). The macroeconomic determinants of volatility in precious metals markets. *Resources Policy*, 35(2), 65–71.
- Bilgin, M. H., Gozgor, G., Lau, C. K. M., & Sheng, X. (2018). The effects of uncertainty measures on the price of gold. *International Review of Financial Analysis*, 58, 1–7.
- Bohl, M. T., Siklos, P. L., & Wellenreuther, C. (2018). Speculative activity and returns volatility of Chinese agricultural commodity futures. *Journal of Asian Economics*, 54, 69–91.
- Cevik, M. S., & Sedik, T. S. (2014). A barrel of oil or a bottle of wine: How do global growth dynamics affect commodity prices? *Journal of Wine Economics*, 9(1), 34–50.
- Chambers, M. J., & Bailey, R. E. (1996). A theory of commodity price fluctuations. *Journal of Political Economy*, 104(5), 924–957.
- Chiang, S. M., Chen, C. D., & Huang, C. M. (2019). Analyzing the impacts of foreign exchange and oil price on biofuel commodity futures. *Journal of International Money and Finance*, 96, 37–48.
- Christoffersen, P. F. (1998). Evaluating interval forecasts. *International Economic Review*, 39(4), 841–862.
- Conrad, C., & Kleen, O. (2019). Two are better than one: Volatility forecasting using multiplicative component GARCH-MIDAS models. *Journal of Applied Econometrics*, 35(1), 19–45.

- Conrad, C., Loch, K., & Rittler, D. (2014). On the macroeconomic determinants of long-term volatilities and correlations in U.S. Stock and crude oil markets. *Journal of Empirical Finance*, 29, 26–40.
- Deaton, A., & Laroque, G. (1996). Competitive storage and commodity price dynamics. *Journal of Political Economy*, 104(5), 896–923.
- Donmez, A., & Magrini, E. (2013). *Agricultural commodity price volatility and its macroeconomic determinants: A GARCH-MIDAS approach*. Joint Research Centre (Seville site). <https://publications.jrc.ec.europa.eu/repository/handle/JRC84138>
- Engle, R. F., Ghysels, E., & Sohn, B. (2013). Stock market volatility and macroeconomic fundamentals. *Review of Economics and Statistics*, 95(3), 776–797.
- Engle, R. F., & Manganelli, S. (2004). CAViaR: Conditional autoregressive value at risk by regression quantiles. *Journal of Business and Economic Statistics*, 22(4), 367–381.
- Fang, L., Chen, B., Yu, H., & Qian, Y. (2018). The importance of global economic policy uncertainty in predicting gold futures market volatility: A garch-midas approach. *Journal of Futures Markets*, 38(3), 413–422.
- Ghysels, E., Sinko, A., & Valkanov, R. (2007). MIDAS regressions: Further results and new directions. *Econometric Reviews*, 26(1), 53–90.
- Gkillas, K., Gupta, R., & Pierdzioch, C. (2020). Forecasting realized gold volatility: Is there a role of geopolitical risks? *Finance Research Letters*, 35, 101280.
- Gleich, B., Achzet, B., Mayer, H., & Rathgeber, A. (2013). An empirical approach to determine specific weights of driving factors for the price of commodities-A contribution to the measurement of the economic scarcity of minerals and metals. *Resources Policy*, 38(3), 350–362.
- Gruber, J. W., & Vigfusson, R. J. (2018). Interest rates and the volatility and correlation of commodity prices. *Macroeconomic Dynamics*, 22(3), 600–619.
- Hamadi, H., Bassil, C., & Nehme, T. (2017). News surprises and volatility spillover among agricultural commodities: The case of corn, wheat, soybean and soybean oil. *Research in International Business and Finance*, 41, 148–157.
- Hansen, B. E. (1994). Autoregressive conditional density estimation. *International Economic Review*, 35(3), 705–730.
- Hu, M., Zhang, D., Ji, Q., & Wei, L. (2020). Macro factors and the realized volatility of commodities: A dynamic network analysis. *Resources Policy*, 68, 101813.
- Hua, R., & Chen, B. (2007). International linkages of the Chinese futures markets. *Applied Financial Economics*, 17(16), 1275–1287.
- Jacks, D. S., & Stuermer, M. (2020). What drives commodity price booms and busts? *Energy Economics*, 85, 104035.
- Jia, R. L., Wang, D. H., Tu, J. Q., & Li, S. P. (2016). Correlation between agricultural markets in dynamic perspective-Evidence from China and the US futures markets. *Physica A: Statistical Mechanics and Its Applications*, 464, 83–92.
- Joets, M., Mignon, V., & Razafindrabe, T. (2017). Does the volatility of commodity prices reflect macroeconomic uncertainty? *Energy Economics*, 68, 313–326.
- Kagraoka, Y. (2016). Common dynamic factors in driving commodity prices: Implications of a generalized dynamic factor model. *Economic Modeling*, 52, 609–617.
- Kilian, L. (2009). Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market. *American Economic Review*, 99(3), 1053–1069.
- Klotz, P., Lin, T. C., & Hsu, S. H. (2014). Global commodity prices, economic activity and monetary policy: The relevance of China. *Resources Policy*, 42, 1–9.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives*, 3(2), 73–84.
- Liu, X. D., Pan, F., Yuan, L., & Chen, Y. W. (2019). The dependence structure between crude oil futures prices and Chinese agricultural commodity futures prices: Measurement based on Markov-switching GRG copula. *Energy*, 182, 999–1012.
- Liu, B. J., Wang, Y., Wang, J., Wu, X., & Zhang, S. (2015). Is China the price taker in soybean futures? *China Agricultural Economic Review*, 7(3), 389–404.



- Mackey, M. C. (1989). Commodity price fluctuations: Price dependent delays and nonlinearities as explanatory factors. *Journal of Economic Theory*, 48(2), 497–509.
- Matesanz, D., Torgler, B., Dabat, G., & Ortega, G. J. (2014). Co-movements in commodity prices: A note based on network analysis. *Agricultural Economics*, 45(S1), 13–21.
- Mo, D., Gupta, R., Li, B., & Singh, T. (2018). The macroeconomic determinants of commodity futures volatility: Evidence from Chinese and Indian markets. *Economic Modeling*, 70, 543–560.
- Natanelov, V., Alam, M. J., McKenzie, A. M., & Van-Huylenbroeck, G. (2011). Is there co-movement of agricultural commodities futures prices and crude oil? *Energy Policy*, 39(9), 4971–4984.
- Ordu, B. M., Oran, A., & Soytaş, U. (2018). Is food financialized? Yes, but only when liquidity is abundant. *Journal of Banking and Finance*, 95, 82–96.
- Prokopczuk, M., Stancu, A., & Symeonidis, L. (2019). The economic drivers of commodity market volatility. *Journal of International Money and Finance*, 98, 102063.
- Tang, K., & Xiong, W. (2012). Index investment and the financialization of commodities. *Financial Analysts Journal*, 68(6), 54–74.
- Theodossiou, P. (1998). Financial data and the skewed generalized T distribution. *Management Science*, 44(12), 1650–1661.
- Tzeng, K. Y., & Shieh, J. C. P. (2016). The transmission from equity markets to commodity markets in crises periods. *Applied Economics*, 48(48), 4666–4689.
- Vercammen, J. (2020). Information-rich wheat markets in the early days of COVID-19. *Canadian Journal of Agricultural Economics*, 68(2), 177–184.
- Wang, X. Y., Luo, Y., Wang, Z. Q., Xu, Y., & Wu, C. X. (2021). The impact of economic policy uncertainty on volatility of China's financial stocks: An empirical analysis. *Finance Research Letters*, 39, 101650.
- Zhang, D., & Ji, Q. (2019). Energy finance: Frontiers and future development. *Energy Economics*, 83, 290–292.
- Zhao, Y., Yang, M., Zhang, Y., & Qi, C. (2010). Impact on the Chinese soybean markets from international prices volatility: Empirical study based on VEC model. *African Journal of Agricultural Research*, 5(15), 1943–1950.