Hindawi
Discrete Dynamics in Nature and Society
Volume 2022, Article ID 1904490, 13 pages
https://doi.org/10.1155/2022/1904490

Research Article

Tail Risk in the Chinese Vegetable Oil Market: Based on the EGAS-EVT Model

Yueqiang Zhang,[1] **Guanghui Han**,[1] **Hui Xie**,[2] **and Zixing Wang**[3]

[1]*School of Management Engineering and Business, Hebei University of Engineering, Handan 056038, China*
[2]*School of Marxism, Hebei University of Engineering, Handan 056038, China*
[3]*Faculty of Social Sciences and Law, University of Bristol, Bristol BS81TH, UK*

Correspondence should be addressed to Guanghui Han; hanguanghui@hebeu.edu.cn

Received 18 November 2021; Revised 21 February 2022; Accepted 1 March 2022; Published 26 April 2022

Academic Editor: Stefan Cristian Gherghina

Copyright © 2022 Yueqiang Zhang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper uses extreme value theory and exponential generalised autoregressive score models to estimate the tail extremes of financial return series. The peak-over-threshold method based on the generalised pareto distribution is combined with the EGAS models and the nonparametric quantile method is used to determine the thresholds in the POT method, which is used to calculate the value-at-risk of financial markets and to perform backtesting. The empirical analysis was conducted on the soybean oil, rapeseed oil, and palm oil futures indices in the Chinese futures market. The study demonstrated that the EGAS-POT models based on nonparametric quantile thresholds can effectively characterise tail risk and provide a feasible measure of risk for investors.

1. Introduction

In recent years, China's economy has made remarkable achievements. At the same time, as the Chinese futures market gradually integrated with the international futures market. In 2020, 6.153 billion lots and 437.53 trillion yuan were traded in China's futures market, accounting for 13.2% of the total volume in the global futures market. Palm oil, soybean oil, and rapeseed oil on China's futures market will rank second, fourth, and ninth, respectively, in the global trading volume of agricultural futures and options. Chinese vegetable oil futures is one of the most widely used varieties in the Chinese market. Moreover, Chinese vegetable oil futures are typical in agricultural futures market. There is a substitute or complementary relationship among all varieties, and it is closely related to spot price. In addition, the price fluctuation of vegetable oils, oilseed, and oilseed meal has a wide range of influence, which has direct or indirect impact on the living consumption of residents, processing of food enterprises, and production of livestock enterprises. Since soybean futures were listed in 1993, oil and oil futures have gradually realized the integration of upstream and downstream of the industrial chain and the diversification of derivatives, such as futures and options. By introducing foreign traders to participate in the trading of palm oil futures and options and other kinds of instruments, they can hedge and avoid risks to a certain extent. China oil futures price is also a leading indicator to monitor the price fluctuation of oil agricultural products and reflect the change of consumer price level. The price fluctuation not only has the transmission mechanism from futures to spot market but also has the cross-market and cross-period relationship. In the context of accelerating the financialisation of agricultural products and preventing systemic financial risks, the role of agricultural futures market in managing price risks is more prominent. Therefore, it is of great theoretical and practical significance to study the risk of vegetable oil futures.

In the context of globalisation, risk measurement has always been an important part of financial risk management and investors and regulators have studied quantitative tools for financial risk, such as value-at-risk [1] (VaR), for this purpose. However, as extreme events have had a huge

impact on financial markets, the volatility in international financial markets has increased, the risk of a global economic downturn is growing, and uncertainty about the outlook has increased dramatically. The tails of traditional metrics describing the distribution of returns are underperforming.

The behaviour of the tails of financial risk has been assessed by many scholars. McNeil [2], Jondeau and Rockinger [3], and Da Silva and Mendez [4] showed that the tails of returns on asset returns have extreme values and do not follow a normal distribution and that their empirical distribution is characterised by spikes and thick tails. Therefore, classical parametric methods based on the assumption of a normal distribution are not suitable for estimating risk in financial markets. One of the ideal alternative parametric methods is extreme value theory (EVT), and methods based on EVT can be used on the basis of VaR. Embrechts [5] discusses the application of EVT to risk modelling. Manfred Gilli and Evis Kellezi [6] apply EVT to measure tail risk in six stock market indices, The results show that extreme value theory is effective for estimating extreme events in financial markets and that the POT method allows better use of information in the data sample to understand the details of financial market data. McNeil [7] showed that direct application of EVT can overestimate or underestimate VaR as the financial asset return series do not satisfy the assumption of independent homogeneous distribution, and there is heteroskedasticity. Therefore, the original income sequence needs to be processed.

Starting from Engle's autoregressive conditional heteroskedasticity [8] (ARCH) model. The problem of constant variance of time series variables in traditional econometrics is solved. More models have been developed to model volatility, such as the generalised autoregressive conditional heteroskedasticity [9] (GARCH) model, the exponential generalised autoregressive conditional heteroskedasticity [10] (EGARCH) model, and the asymmetric power generalised autoregressive conditional heteroskedasticity [11] (APARCH) model. These models perfectly interpret the volatility characteristics of time series in financial markets, such as asymmetry and leverage. As financial markets become more complex. Creal, Koopman, and Lucas [12] creatively proposed a unified framework for modelling time-varying parameters, namely the generalised autoregressive score (GAS) models, which provides a new option for modelling financial asset return volatility. Nortey et al. [13] modelled the extreme values of stock index volatility in Ghana by the autocorrelation of the return series was corrected and the conditional heteroskedasticity in the presence of collections. Applying EVT to fit the tails of daily stock return data for Ghana, the study showed that among the methods used to estimate the parameters, the maximum likelihood estimation (MLE) method provides more accurate estimates. Taking the top 10 sector indices of the SSE as an example, Ping [14] developed a GAS volatility model to predict and compare the out-of-sample VaR forecasting effects of the model. The study showed that the GAS volatility model for time-varying volatility modelling can effectively use the valid information of the distribution, and VaR performs better. To analyse the correlation between different types in the financial market, Rongda and Jianjun [15] developed a multivariate GAS model to analyse the interaction between the dependence balances and volatility between the prices of crude oil and gold, and the results show that the predictive power of volatility and correlation in the multivariate GAS model outperforms the DCC-GARCH model. Lazar and Xiaohan [16] introduced intraday information into the GAS model in quantile regression setting to estimate risk. The results show that the GAS model, augmented by the implemented volatility metric, consistently outperforms other models across all indices and various probability levels.

Traditional time-series models, such as AR model, MA model and GARCH family model, were mostly used in previous studies. In addition, they tend to predict the return series or measure the risk of the return series of stock index. There are few researches on the risk measure of the return series of agricultural futures and futures index in the financial market. In this paper, the time-varying parameters based on the score function are introduced into the EGARCH model for the futures index of Chinese vegetable oils and fats, and the most suitable residual distribution is selected. A nonparametric method is proposed to select the threshold value for the extreme value of the filtered standardized residual sequence, and the EGAS-POT model is established by combining the extreme value theory. Compared with the traditional threshold selection method, the utility of the model in risk measurement is investigated and tested back. Based on the empirical analysis results, some reasonable suggestions are put forward for the risk management of vegetable oil futures in China.

2. Materials and Methods

2.1. EGAS Volatility Model. The GAS model also goes by the name of dynamic conditional score (DCS) model, Score driven (SD) model, or dynamic score (DySco) model, is a time-varying volatile parametric models driven by observations that allows the model parameters to vary as the score function of the log-likelihood function changes. The dynamic behaviour of the time-series process is portrayed through the dynamics of the parametric variables leading to the variables and exogenous variables. The EGAS model, on the contrary, builds on the GAS model using the logarithm of the conditional variance instead of the conditional variance, allowing for asymmetries in positive and negative asset returns on volatility, thus allowing the dynamics of the impact of positive and negative returns on volatility to be captured effectively.

Assuming that y_t is the financial time-series observation, σ_t is the time-varying conditional y standard deviation, which represents the volatility of the time-series data, and F^{t-1} is the information set at moment $t-1$, then the observation y_t probability density function is as follows:

$$p\left(y_t|\sigma_t, F^{t-1}\right). \tag{1}$$

Then the expression for the EGAS model based on time-varying volatile is as follows:

$$\begin{cases} y_t = \mu + \sigma_t z_t, \\ \ln \sigma_{t+1}^2 = \omega + \sum_{i=1}^{p} A_i s_t + \sum_{j=1}^{q} B_i \ln \sigma_t^2, \nabla_t = \dfrac{\partial \ln \ p\left(y_t|\sigma_t, F^{t-1}\right)}{\partial \ln \ s_t}, \end{cases} \begin{matrix} s_t = S_{t-1}\nabla_t, \\ S_t = I_t^{-\gamma}, \end{matrix} \quad I_t = \left[\nabla_t \nabla_t'\right]. \tag{2}$$

where when $t = 1$, σ_1 is the unconditional standard deviation, z_t is the standardized residual series, A_i and B_i $(i = 1, 2, \ldots, p,\ j = 1, 2, \ldots, q)$ are time-varying coefficient matrices, reflecting the time-varying nature of the fluctuations and the aggregation and mean recovery of the fluctuations, respectively, usually p and q can be taken as 1. ω is the constant vector; I_t is the information matrix; S_t is the deflation matrix; in general, γ takes 0, at this time; S_t is the deflation matrix is the unit deflation matrix; ∇_t is the score function corresponding to σ_t, is the core driving term of the EGAS fluctuation model.

When the standardized residuals z_t obey a different distribution, the expression for the score function changes as well. For financial time series, the distributions often assumed are: the Gaussian (Normal) distribution (N), the standard student T distribution (ST), the generalised error distribution (GED), and the skewed student T distribution (SKST), where the probability density of the skewed student T distribution is as follows:

$$f(z) = bc\left[1 + \frac{1}{v-2}\left(\frac{bz + a}{1 + \operatorname{sgn}(bz + a)l}\right)^2\right]^{-v+1/2}, \tag{3}$$

where $c = 1/[\sqrt{v-2}\,B(1/2, v/2)]$, $a = 4\lambda c(v-2)/(v-1)$, $b^2 = 1 + 3\lambda^2 - a^2$, $B(\cdot)$ is the beta function, $\operatorname{sgn}(\cdot)$ is the symbolic function, and z is the variable with the mean value of 0 and the variance of 1. Parameter v represents the kurtosis of the skew t distribution. When v is smaller, the kurtosis is larger, and the thick tail of the spike is more obvious. λ is an asymmetric coefficient, indicating the skewness of the skew t distribution. If $\lambda > 0$, the distribution is right, and if $\lambda < 0$, the distribution is left. The skew t distribution includes normal distribution, skew normal distribution, and t distribution. When the parameter $v \longrightarrow \infty$ and $\lambda = 0$, the distribution is normal distribution. When parameter $v \longrightarrow \infty$. At this time, the distribution is skew normal distribution. When the parameter $\lambda = 0$, the distribution is t.

When the standardized residuals obey SKST, the score function is as follows:

$$\nabla_t = \frac{\partial \ln \ p\left(y_t|\sigma_t, F^{t-1}\right)}{\partial \ln \ \sigma_t} = \frac{bz_t(v+1)(bz_t + a)}{(bz_t + a)^2 + v[1 + \operatorname{sgn}(bz_t + a)\lambda]^2} - 1. \tag{4}$$

As can be seen from the score function, the skewness parameter v and the kurtosis parameter λ determine the value of the score function, and as the value of v increases, the score function becomes more sensitive to extreme values; the skewness parameter λ reflects the sensitivity of the score function to shocks on the left side, when the distribution is left-skewed, the score is relatively more sensitive to shocks on the right side; when the distribution is right-skewed, the score is relatively more sensitive to shocks on the left side; when the distribution is symmetrical, the response of the score to shocks on both sides is symmetric.

2.2. Extreme Value Theory. Extreme value theory was introduced by Gnedenk [17]. Also known as the law of small numbers is primarily concerned with the prediction of rare events. The theory aims to investigate the distribution of the extremes of a sequence, using the generalised pareto distribution or generalised extreme value distribution (GEV) to approximate the tail distribution of losses. First applied in hydrology, seismology, and climatology, it is commonly used to analyse probabilistic rare cases. With the increasing refinement of the theory, EVT research has been applied to science and technology, engineering and other fields, with Longin [18] pioneering the use of EVT in risk management with good results. In financial engineering, for the tail characteristics of risk loss distribution, it is usually used to analyse events with rare probability. It can rely on a small amount of sample data to obtain the change of extreme value in the overall distribution when the overall distribution is unknown and has the ability to estimate beyond the sample data.

There are two methods for the application of extreme value theory: Block maxima (BM) method and peak-over-threshold (POT) method, as the use of the BM method leads to the absence of extreme data in the block and the loss of extremely valuable extreme information. This paper uses the POT method, which is more widely used in practice. Clément Dombry and Ana Ferreira [19] show that the POT method is preferable when considering MLE, and the estimation results are more convincing for extreme data values.

Assume that the data of the random variable sequence is $x_1, x_2, \ldots, x_n$ is independently and identically distributed, and the distribution function of the random variable is $F(x)$, let x_m be the maximum value of the random variable sequence data, by setting the threshold value $u\,(u < x_m)$, and all the observed data above this threshold value form a data group $\{Z_i\}$, with this data group as the object of modelling, such that $K_i = Z_i - u$, then we have

$$F_u(k) = P(Z - u \le k | X > u), \quad k \ge 0. \tag{5}$$

The derivation of the conditional probability formula leads to the following:

$$\begin{aligned} F_u(k) &= \frac{F(u+k) - F(u)}{1 - F(u)} \\ &= \frac{F(Z) - F(u)}{1 - F(u)} \Rightarrow F(Z) \\ &= F_u(k)[1 - F(u)] + F(u). \end{aligned} \tag{6}$$

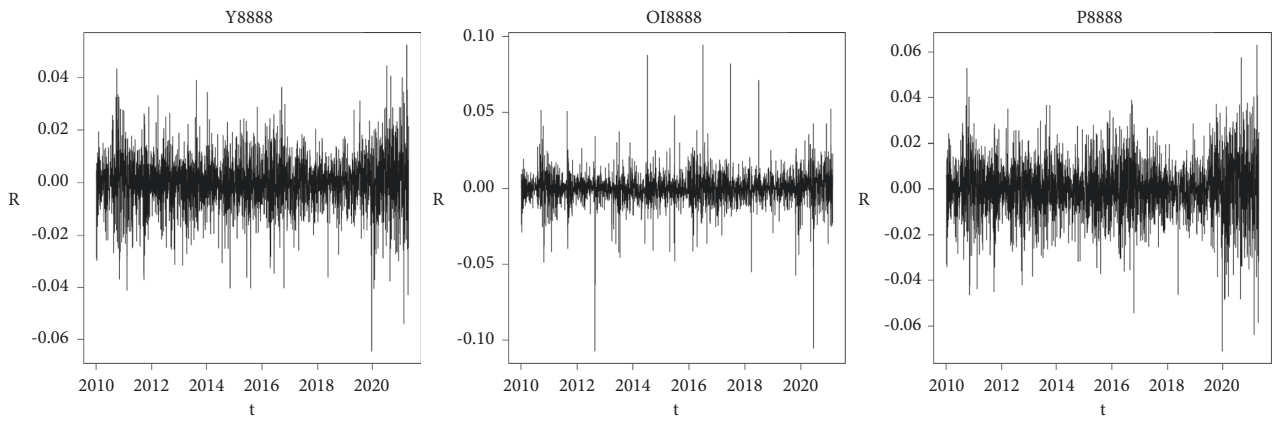

Figure 1: Daily returns trend.

Table 1: Daily returns observations descriptive statistics.

Statistical characteristics	Soybean oil	Rapeseed oil	Palm oil
Sample size	2767	2728	2767
Minimum value	−0.06447	−0.10728	−0.07128
Maximum value	0.05254	0.094483	0.06311
Mean value	1.922*e*-005	7.0085*e*-005	9.4054*e*-006
Standard deviation	0.010503	0.010719	0.012617
Skewness	−0.17887	−0.21697	−0.20044
Super kurtosis	2.1892	15.5152	1.9282

When the threshold u is taken to be relatively high, the suprathreshold distribution will converge to the GPD. The GPD expression is the distribution function $F_u(y)$ approximating the $G'_{\xi,\eta}(y)$ generalised Pareto distribution.

$$\underset{u \longrightarrow \infty}{F_u(k)} \approx G'_{\xi,\eta}(k) = \begin{cases} 1 - \left(1 + \dfrac{\xi}{\eta} y\right)^{-1/\xi}, & \xi \neq 0, \\ 1 - \eta^{-k/\eta}, & \xi = 0, \end{cases} \tag{7}$$

where ξ and η are the shape and scale parameters, respectively. When $\xi \geq 0$, $y \geq 0$, indicating a thick tail of the distribution function, and the presence of extreme values; when $\xi < 0$, $y \in [0, -\xi/\eta]$. The probability density function of the GPD is known, and hence the log likelihood function of a sequence of random variables known to obey an independent distribution:

Table 2: Jarque-Bera test for daily returns.

	X-squared	P value
Soybean oil	567.31	6.4547*e*−124
Rapeseed oil	27,382.00	0
Palm oil	447.19	7.8144*e*−98

$$L(\xi, \eta | k) = \begin{cases} -n \ln \eta - \left(1 + \dfrac{1}{\eta} k\right) \displaystyle\sum_{i=1}^{n} \ln\left(1 + \dfrac{\eta}{\xi} k_i\right), & \xi \neq 0, \\ -n \ln \eta - \dfrac{1}{\eta} \displaystyle\sum_{i=1}^{n} k_i, & \xi = 0. \end{cases} \tag{8}$$

When u is determined, the estimates of ξ and η are obtained by MLE according to equation (8), and the shape parameter ξ reflects the tail of the distribution. At the same time, the number of observations of the random variable series data that exceed the threshold u data can be obtained, denoted as N_u, and the new expression can be obtained by replacing the value of $F(u)$ with the frequency according to equation (6):

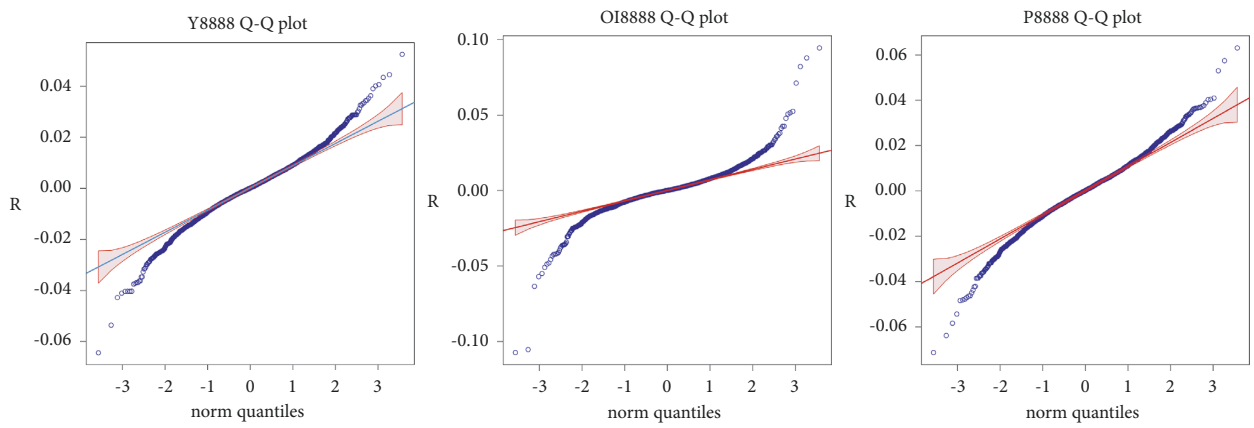

Figure 2: Normal Q-Q plots of daily returns.

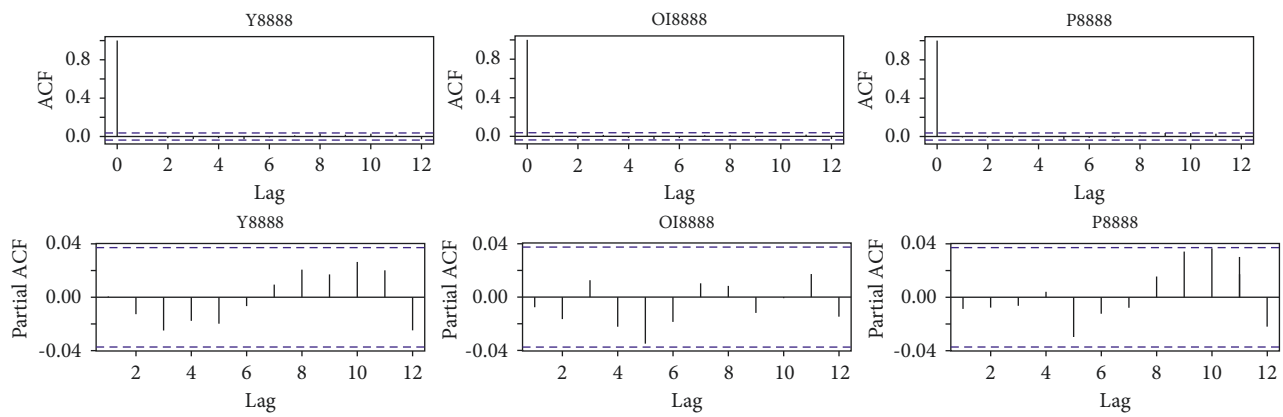

Figure 3: Autocorrelation and partial autocorrelation of daily returns.

$$F_u(k)[1-F(u)]+F(u)=\begin{cases}\dfrac{N_u}{N}\left\{1-\left[1+\dfrac{\xi}{\eta}(z-u)^{-1/\xi}\right]\right\}+\left(1-\dfrac{N_u}{N}\right),\\[2ex] \dfrac{N_u}{N}\left[1-\eta^{-(z-u)/\eta}\right]+\left(1-\dfrac{N_u}{N}\right),\end{cases}$$
$$=\begin{cases}1-\left[1+\dfrac{\xi}{\eta}(z-u)\right]^{-1/\xi}, & \xi\neq 0,\\[2ex] 1-\dfrac{N_u}{N}\eta^{-(z-u)/\eta}, & \xi=0.\end{cases} \tag{9}$$

2.3. VaR Estimation Based on the POT Method. VaR is the maximum possible loss to an investor owning a single asset or portfolio of assets at a certain confidence level p (99%, 95%) and holding period. Its essence is to calculate the tail quantile of the yield distribution, where the long VaR corresponds to the lower tail quantile of the yield distribution and the short VaR corresponds to the upper tail quantile of the yield distribution. The expression of VaR_p is

$$VaR_p=-\inf\{|x|f(X\leq x)>(1-p)\}. \tag{10}$$

Taking equation (9) into equation (10) gives the following:

$$VaR_p = \begin{cases} u - \frac{\eta}{\xi}\left[1 - [-n\ln(1-p)]^{-\xi}\right], & \xi \neq 0, \\ u - \eta\ln[-n\ln(1-p)], & \xi = 0. \end{cases} \quad (11)$$

3. Empirical Analysis

3.1. Data Analysis. The empirical part uses the Chinese futures market Soybean oil (Y8888), Rapeseed oil (OI8888), and Palm oil (P8888) futures indices as raw data (data source: Flush iFinD), with the sample space selected from 4 January 2010 to 31 May 2021, and the first-order difference of the logarithm of the daily closing price is used as the daily log return for ease of processing, that is, $R_t = LnP_t - LnP_{t-1}$, where P_t denotes the closing price on day t and P_{t-1} denotes the closing price on day $t-1$. Figure 1 shows a graph of the daily return series.

Descriptive statistics for the daily return series are shown in Table 1.

As shown in Table 1, from the description of daily logarithmic returns, the average returns of soybean oil, rapeseed oil, and palm oil futures are all near 0, with a range of 0.117006, 0.201763, and 0.134392, respectively. The excess kurtosis coefficient is greater than 0, that is, the logarithmic rate of return series has a peak. The skewness is less than 0, indicating that there are different degrees of left skewness, indicating that there are more huge falls in the market than huge rises, that is, there is a negative skewness vegetable oil futures all exhibit a left-skewed and spiky distribution that does not obey a normal distribution. This is the same result as Balaban's [20] study on the distribution characteristics of daily stock returns and their asymmetry. The test used to test the daily returns for smoothness is the Jarque-Bera [21] test, and the results are shown in Table 2. The original hypothesis is rejected because the significance is much less than the critical value of its significance level of 1%, and the series does not have a unit root, and is a smooth series.

In order to further test the distribution characteristics of the sample series, Figure 2: normal Q-Q plots of daily returns was described. The scattered points on the log-return normal Q-Q plots were curved at both ends and distributed outside the 95% confidence level interval of the normal distribution, indicating that the distribution of log-return is thick-tailed. To sum up, the original sample sequence follows the distribution with sharp peak and thick tail deviating to the left, so the predicted results of VaR calculation method based on normal distribution are too conservative.

Second, part of the efficiency of financial markets also reflects the general autocorrelation between raw returns, so using metrics directly is not feasible. Figure 3: ACF plot shows that there is no autocorrelation in the daily return series, and the autoregressive conditional heteroskedasticity test on the return residual series shows that there is a strong ARCH effect in the residual series through the ARCH-LM test. The POT method requires the sequence of random variables to meet the requirement of independent identical distribution, so for the original data with volatility aggregation and leverage effect, so the volatility model needs to be constructed to filter the return series, and the residual series of each set of returns obeying independent identical distribution is found.

Table 3: EGAS (1,1) model AIC and SC estimation results.

	Assumed distribution	AIC	SC
Soybean oil	N	−6.3496	−6.3498
	ST	−6.3562	−6.3937<
	GED	−6.3994	−6.3908
	SKST	−6.4038<	−6.3931
Rapeseed oil	N	−6.2403	−6.2338
	ST	−6.6442<	−6.6355<
	GED	−6.620	−6.6121
	SKST	−6.6436	−6.6327
Palm oil	N	−6.0132	−6.0068
	ST	−6.0457	−6.0372<
	GED	−6.0421	−6.0335
	SKST	−6.0472<	−6.0365

Table 4: Results of parameter estimation for EGAS (1,1)-SKST model.

	A_1	B_1	λ	ν
Soybean oil	0.051642	0.990683	−0.061800	6.800666
Rapeseed oil	0.118198	0.97477	−0.010878	3.297259
Palm oil	0.047381	0.994420	−0.063079	7.926744

Table 5: Thresholds **u** selection.

	Soybean oil	Rapeseed oil	Palm oil
Long position u	−1.279089	−1.070747	−1.521152
Short position u	1.219153	1.048097	1.188821

Table 6: GPD fitting parameters.

	Long position		Short position	
	ξ	η	ξ	η
Soybean oil	0.5710	0.0528	0.2650	0.5560
Rapeseed oil	0.4980	0.3090	0.0386	0.7820
Palm oil	0.0232	0.8240	0.0269	0.5500

3.2. Model Parameter Estimation. In this study, the EGAS (1,1) model was chosen to filter the data for each set of return series and to compare the fit of the model under various hypothetical distributions according to the AIC criterion and the SC criterion. The AIC and SC estimation results of the model are shown in Table 3:

As can be seen from Table 3, the hypothetical distribution with skewed, spiky and thick-tailed characteristics is significantly better than the symmetric hypothetical distribution, and the hypothetical distribution of raw returns is chosen as SKST.

According to the parameter estimation results in Table 4, the parameter A_1 was significantly greater than 0, indicating that the return rate series had obvious time-varying

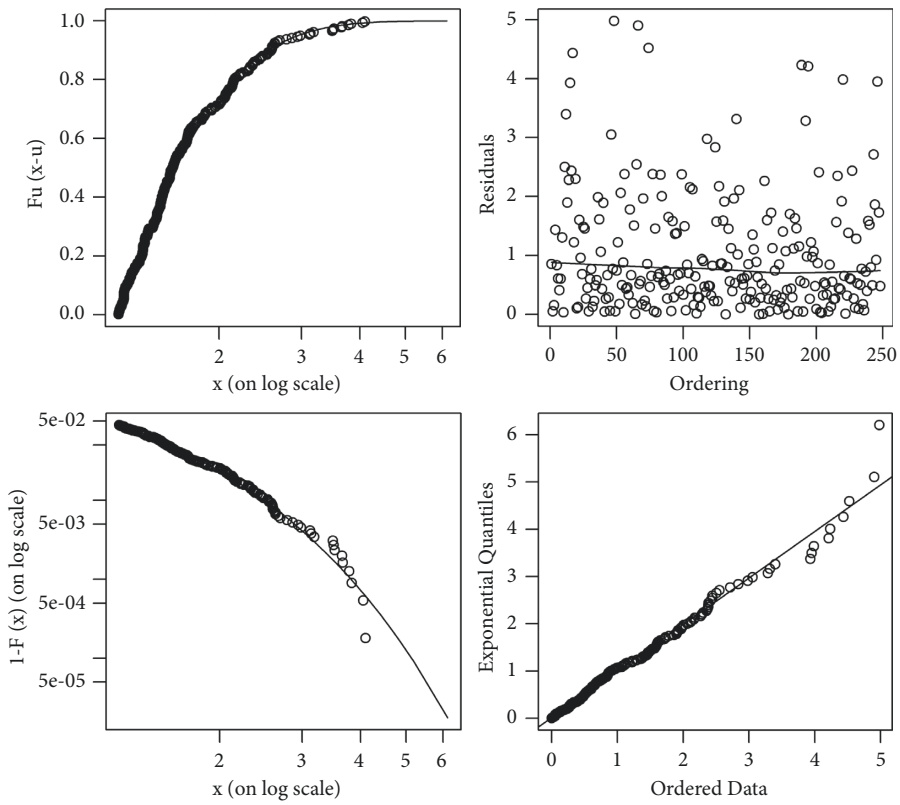

Figure 4: Soybean oil positive returns GPD fitted diagnostic plots.

fluctuation characteristics. Parameter B_1 is close to 1, which means that the return rate series has a strong agglomeration, $A_1 < B_1$, which means that the unexpected news impacts the fluctuation of return, and $A_1 > 0$, it proves that this kind of shock is positive, that is, the occurrence of fluctuation is usually followed by a larger fluctuation in the later period. In terms of distribution parameters, the palm oil futures parameter ν is larger, reflecting more extreme risk exposure in historical data.

3.3. The POT Method. The residual series obtained by constructing the EGAS (1,1)-SKST model satisfies the POT method requirements such that the residual series $\{X_t\} = \{z_t\}$ and the distribution function is fitted with GPD model for data above the threshold extremes. In performing model estimation, the upper and lower tail thresholds of the residual series are first determined. Caeiro [22] used different methods for the selection of the thresholds a comparative analysis was carried out.

In this paper, we adopt a novel nonparametric method for selecting the threshold u. Applying Grevenko's theorem [23], the empirical distribution function is related to the overall distribution function, thereby improving the accuracy of the quantile of the original data series.

Step 1. If the number of data in the residual series is N, let the parameter $M = 100$, extract 100 data item by item in the residual series in time order, select the 2nd to 101 data for the 2nd time, ..., select the ith to $i + 99$ data for the ith time, and so on, to obtain $(N - M + 1)$ data.

Step 2. Based on the confidence level α, the quantile values of each group of data were calculated separately using the historical simulation method combined with the principle of Mouchel [24] and through Holger Drees [25] in a comparative study of random and deterministic thresholds, so that $\alpha = 10\%$ to obtain $(N - M + 1)$ quantile values.

Step 3. The average of the selected quantiles is noted as s. The value closest to s is found in the new interest sequence, which is noted as the threshold value u, and u is used as the threshold value in the POT method.

The nonparametric method not only effectively avoids the subjective judgement of thresholds based on image methods that lead to over or improper fitting, but for the value of α avoids the situation in the POT method where extreme values are piled up leading to too large a choice of thresholds and a small sample of extreme value data. The results based on the nonparametric threshold selection are shown in Table 5.

The results of the nonparametric method of fitting are shown in Table 6.

In order to test the fitting effect of the POT method, we further give Figures 4 to 9 GPD fitting diagnostic plots for the fitting of the residual series. Observe that the data points in the graph are concentrated in each distribution curve except for individual data, proving that the POT method fits

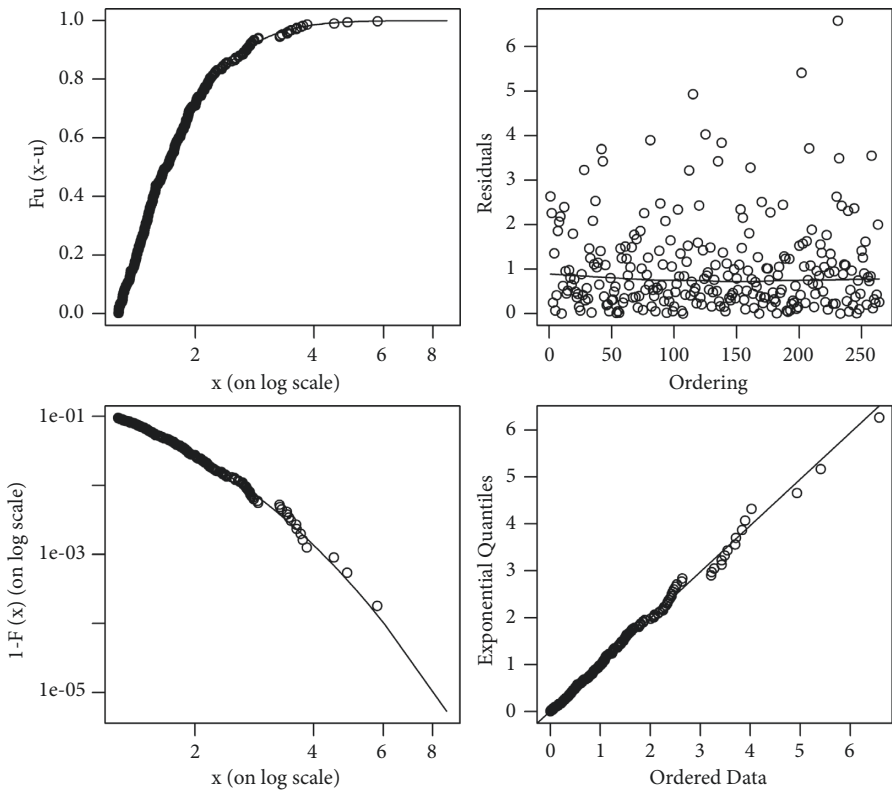

Figure 5: Soybean oil negative returns GPD fitted diagnostic plots.

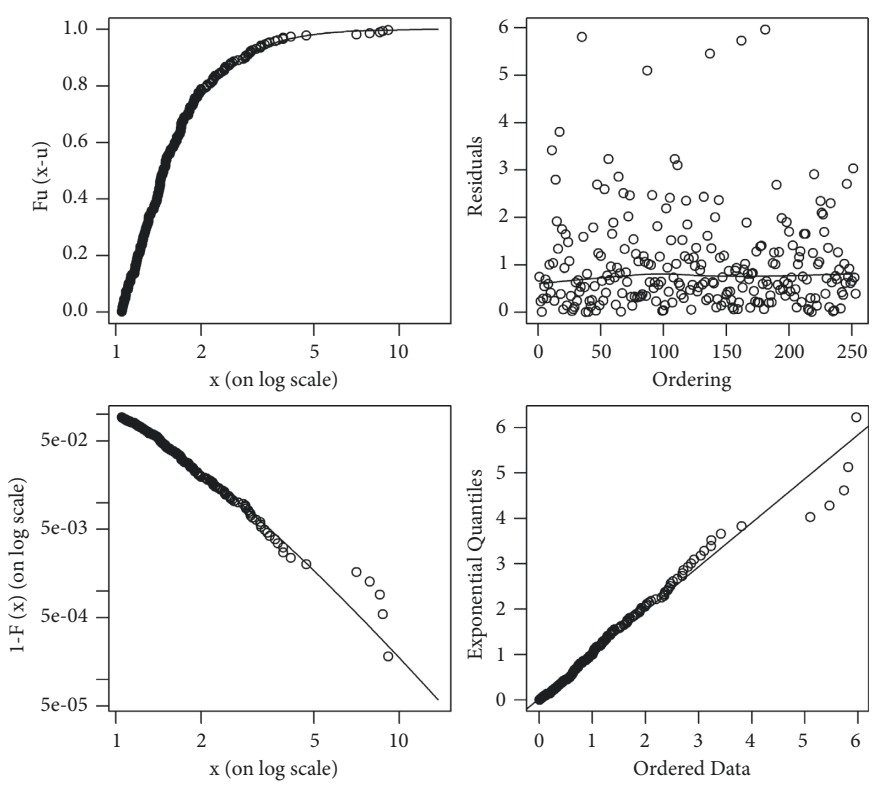

Figure 6: Rapeseed oil positive returns GPD fitted diagnostic plots.

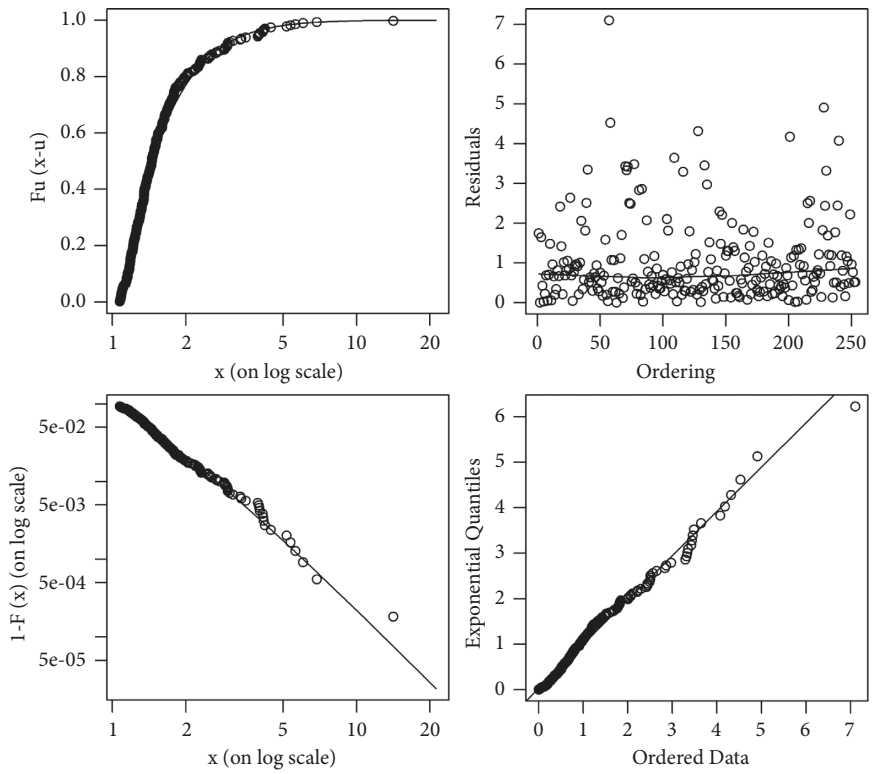

Figure 7: Rapeseed oil negative returns GPD fitted diagnostic plots.

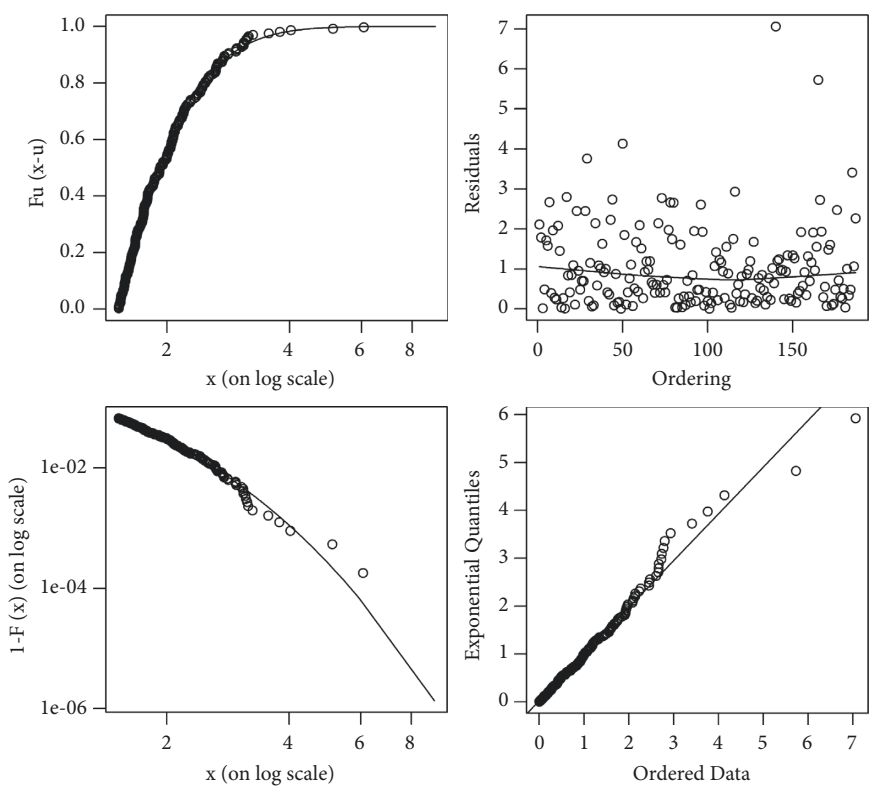

Figure 8: Palm oil positive returns GPD fitted diagnostic plots.

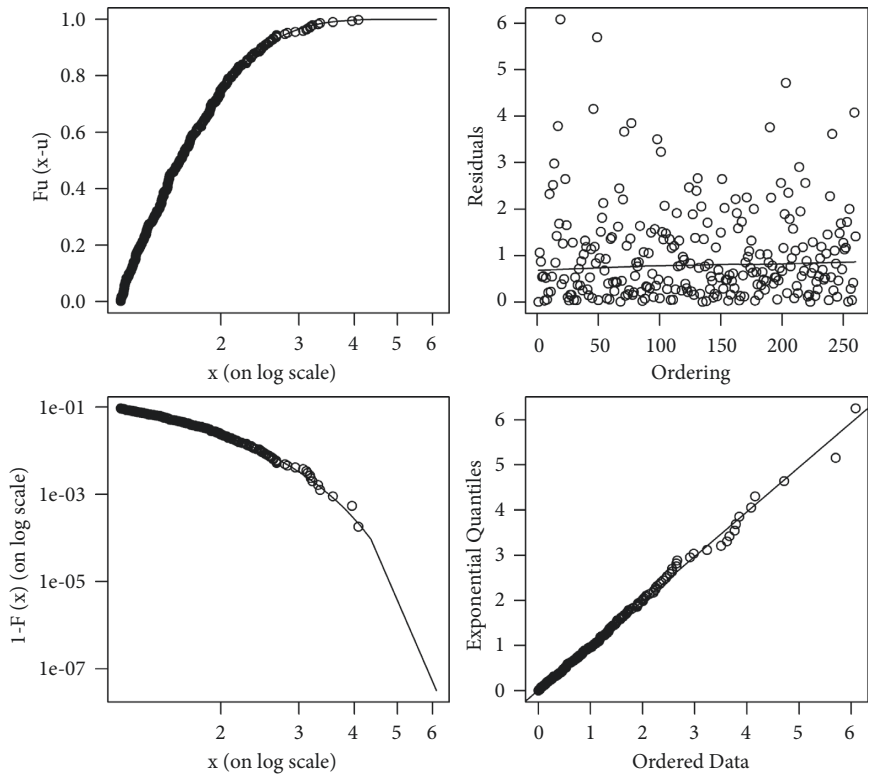

Figure 9: Palm oil negative returns GPD fitted diagnostic plots.

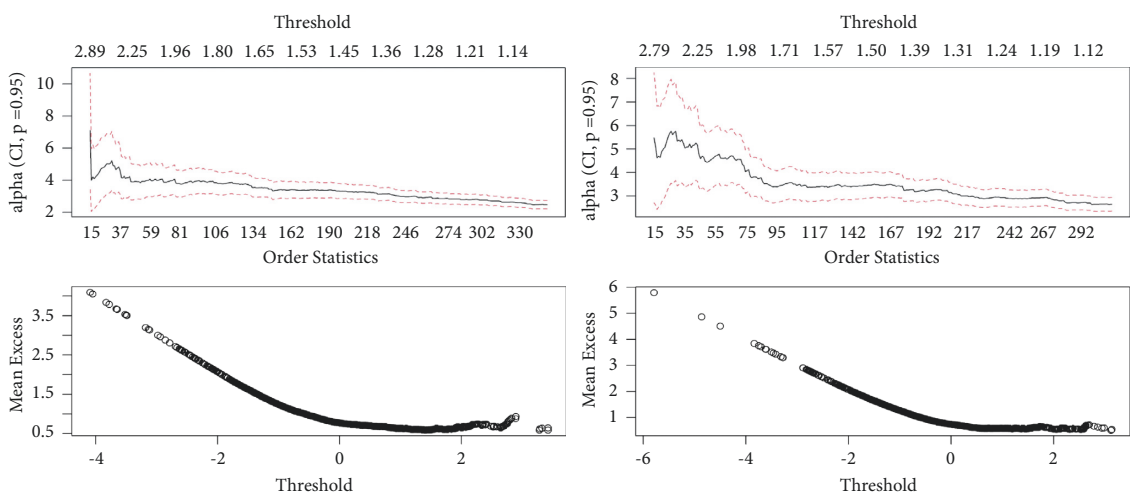

Figure 10: Soybean oil-Hill plots and the MEF plots. The top is the Hill plot, the bottom is the MEF plot, the left is the negative returns plots and the right is the positive returns plots.

the extreme data of the new coupon series well and can accurately reflect the tail characteristics of the true loss distribution of the extreme risk of the returns series.

In order to compare the validity of the nonparametric method of threshold selection, the Hill estimator [26] and the mean excess function [27] (MEF) were used to estimate the VaR values by Figure 10 selecting the threshold u_1 and performing a comparative analysis. The Hill plots Figure 10 and MEF plots are shown in Figures 10 to 12.

The thresholds u_1 were selected by observing Hill plots and MEF plots, and the results of the threshold selection are shown in Table 7.

Define the nonparametric method to select the threshold value of VaR is VaRn, Hill estimation method and MEF method to select the threshold value of VaR is VaRm, respectively, at the confidence level p of 99% and 95% of the VaR value, The estimation results are shown in Tables 8 and 9.

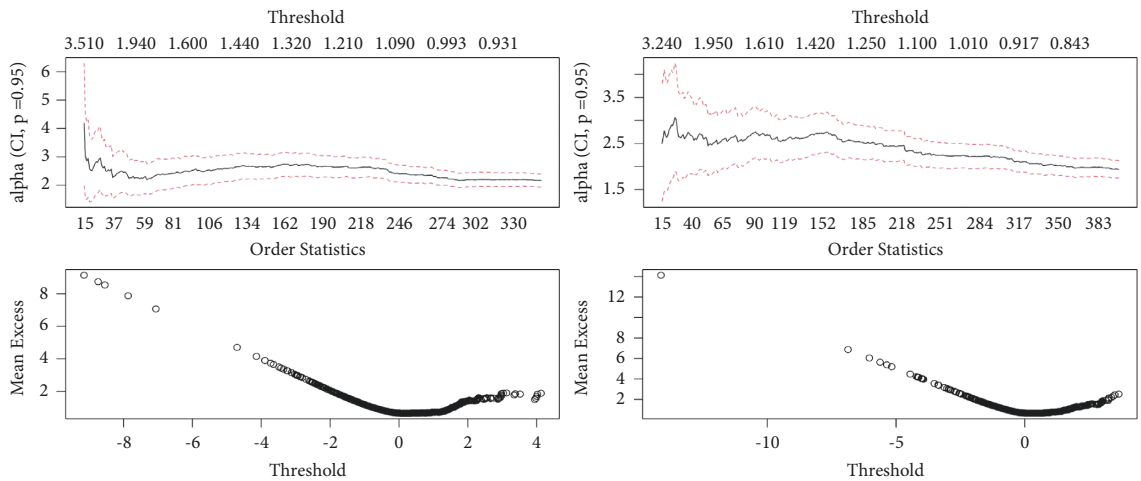

Figure 11: Rapeseed oil-Hill plots and the MEF plots.

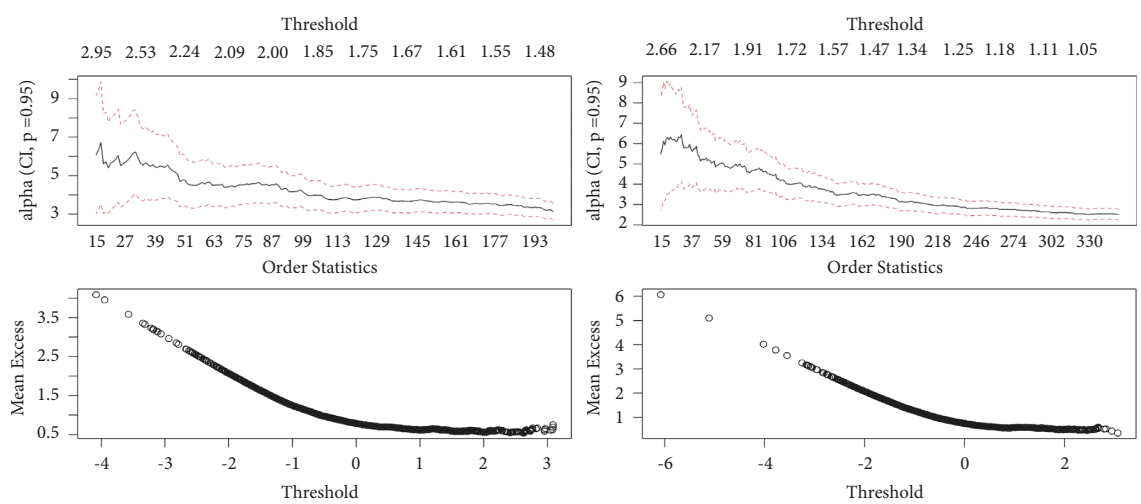

Figure 12: Palm oil-Hill plots and the MEF plots.

Table 7: Thresholds $\mathbf{u}_1$ selection.

	Soybean oil	Rapeseed oil	Palm oil
Long position u_1	−1.280722	−1.007843	−1.590512
Short position u_1	1.213788	1.033504	1.161942

Table 8: VaRn estimation results.

	p	Long position $\mathrm{VaR_n}$	Short position $\mathrm{VaR_n}$
Soybean oil	95%	−1.640087	1.556682
	99%	−2.664474	2.502442
Rapeseed oil	95%	−1.411081	1.421339
	99%	−2.677870	2.762528
Palm oil	95%	−2.063909	2.015324
	99%	−3.435836	3.185844

3.4. VaR Backtesting. The test for VaR takes the Kupiec test [28], which is a very widely used method of posterior analysis, by constructing a likelihood ratio (LR) statistic to test the estimated loss value and the actual loss value, which passes the test within a certain acceptance range, and under the original hypothesis, the statistic LR obeys a χ^2 distribution with degree of freedom of 1. The smaller the statistic LR, the larger the P value, indicating that in the model, the more accurate and the higher the credibility. When $P \geq 0.05$, its validity passed the post hoc test.

Table 9: $\mathrm{VaR_m}$ estimation results.

	p	Long position $\mathrm{VaR_m}$	Short position $\mathrm{VaR_m}$
Soybean oil	95%	−1.631120	1.556638
	99%	−2.643761	2.501126
Rapeseed oil	95%	−1.386436	1.394525
	99%	−2.696885	2.759419
Palm oil	95%	−1.641851	1.580744
	99%	−2.667628	2.471242

$$\mathrm{LR} = 2\ln\left[\left(1-\frac{N_u}{N}\right)^{N-N_u}\left(\frac{N_u}{N}\right)^{N_u}\right] - 2\ln\left[(1-p)^{N-N_u}p^{N_u}\right], \quad (12)$$

Table 10: VaRn backtesting results.

	Soybean oil		Rapeseed oil		Palm oil	
Long position VaRn						
p	99%	95%	99%	95%	99%	95%
N_u	33	138	32	146	27	148
LR	0.97685	0.00093	0.78150	0.69594	0.01651	0.69344
P	0.32297	0.97564	0.37668	0.40414	0.89773	0.40499
Short position VaRn						
p	99%	95%	99%	95%	99%	95%
N_u	37	160	29	158	27	142
LR	2.87393	3.40267	0.10733	3.43356	0.01651	0.10053
P	0.09002	0.06509	0.74320	0.06388	0.89773	75119

Table 11: $\mathrm{VaR_m}$ backtesting results.

	Soybean oil		Rapeseed oil		Palm oil	
Short position VaRm						
p	99%	95%	99%	95%	99%	95%
N_u	34	140	32	151	59	225
LR	1.36344	0.02063	0.78150	1.59221	27.0479	inf
P	0.24294	0.88577	0.37668	0.20701	$1.98E{-}07$	0
Short position VaRm						
p	99%	95%	99%	95%	99%	95%
N_u	37	160	29	165	80	239
LR	2.87393	3.40268	0.10733	5.93351	66.2146	inf
P	0.09002	0.06509	0.74320	0.01486	$4.04E{-}16$	0

where N_u is the number of days to failure and N is the total number of days observed. N_u/N is the frequency of failure and p is the confidence level.

The backtesting for the raw returns of the three futures indices are presented in Tables 10 and 11 for the long and short positions under the two threshold selection methods, respectively.

From Tables 10 and 11, it can be seen that the palm oil futures index is exposed to greater risk than the other two futures indices at the 99% and 95% confidence levels. The reason is that the gap between China's production and demand ranks among the top in the world, and China is excessively dependent on imports of palm oil. The EGAS-SKST-POT model under the threshold selected using the nonparametric method, the number of days to failure is closer to the theoretical number of days and the VaR values of long and short positions under the model pass the model backtesting, proving that the model is feasible. The use of this approach enables a significant increase in the accuracy of the model, and out-of-sample forecasting ability. In contrast, the thresholds based on the Hill estimator and the empirical mean-excess function deviate significantly and the VaR values for long and short positions under the model do not all pass the model backtesting.

4. Conclusion

The main objective of this paper is to apply EVT and the EGAS model to the Chinese vegetable oil futures index by targeting the characteristics of aggregation, persistence and asymmetry of daily returns in the Chinese stock market. First, the standard residual series based on the SKST distribution is inscribed through the EGAS(1,1) model, and a nonparametric quantile-based approach is adopted to select the threshold and apply the POT method in extreme value theory to calculate VaR values and perform backtesting. The study shows that the nonparametric method proposed in the article is able to select suitable thresholds and that suitable estimates can be obtained by fitting the GPD distribution for data where the new series exceeds the threshold, as well as the feasibility of the EGAS model on the China Futures Index.

Based on the empirical analysis of the value at risk of soybean oil, rapeseed oil, and palm oil futures indexes, the following suggestions are put forward:

First of all, in order to prevent the price fluctuation and risk of vegetable oil futures market, we should give full play to the price discovery and hedging functions of futures market. In the current situation of increasing price volatility in the futures market, it is necessary to strengthen effective monitoring of vegetable oil futures prices, improve the financial market supervision system, and prevent abnormal price volatility from negatively affecting other futures varieties. By monitoring and real-time analysis of abnormal events such as agricultural futures market information and other market information and policy changes, effective supervision of risk events can be achieved, and then timely warning can be given before the risk may break out, and timely response can be made after the risk occurs.

Second, strengthen the construction of futures market, vigorously develop futures market, increase trading varieties, control prices, give play to the unity and advance advantage of futures and options market information, and develop in the direction favorable to the development of financial market. Establish cross-sector agricultural product market risk early warning organization coordination mechanism. To solve the problems of incomplete and asymmetric information of agricultural products market, we should construct a multilevel matching and linkage information system of agricultural products market. Build agricultural market information sharing platform, form multidepartment organization and coordination mechanism.

Finally, as a result of the palm oil futures price volatility compared with other oil futures price volatility is larger, the sino-us trade friction and imposing import tariffs under the background of U.S. soybeans, necessary policy interventions can be taken to reduce oil futures price volatility, focus on palm oil futures varieties, broaden the import channel, implementing multiple imports of palm oil, add palm oil strategic reserve. Participants and investors in relevant industries should enhance policy attention and market sensitivity, comprehensively consider the factors of market fluctuations, and further strengthen rational production and investment awareness. In addition, the government should improve laws and regulations related to palm oil production and trade and reasonable market structure to ensure the safety and virtuous cycle of vegetable oil market and related markets.

Data Availability

The data used to support the results of this study are available from the Flush iFinD.

Conflicts of Interest

The authors declare that they have no conflicts of interest related to this work.

References

[1] J. P. Morgan, "Riskmetrics technical document," New York, NY, USA, 1996.

[2] A. J. McNeil, "Estimating the tails of loss severity distributions using extreme value theory," *ASTIN Bulletin*, vol. 27, no. 1, pp. 117–137, 1997.

[3] E. Jondeau and M. Rockinger, "The tail behavior of stock returns: emerging versus mature markets," *HEC Paris*, vol. 66, 1999.

[4] A. Carvalhal and B. V. Mendes, "Value-at-risk and extreme returns in Asian stock markets," *International Journal of Business*, vol. 8, no. 1, 2003.

[5] P. Embrechts, C. Klüppelberg, and T. Mikosch, "Modelling extremal events: for insurance and finance," *Springer Science & Business Media*, vol. 33, 2013.

[6] M. Gilli, "An application of extreme value theory for measuring financial risk," *Computational Economics*, vol. 27, no. 2, pp. 207–228, 2006.

[7] A. J. McNeil and R. Frey, "Estimation of tail-related risk measures for heteroscedastic financial time series: an extreme value approach," *Journal of Empirical Finance*, vol. 7, no. 3-4, pp. 271–300, 2000.

[8] R. F. Engle, "Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation," *Econometrica*, vol. 50, no. 4, pp. 987–1007, 1982.

[9] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of Econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[10] D. B. Nelson, "Conditional heteroskedasticity in asset returns: a new approach," *Econometrica*, vol. 59, no. 2, pp. 347–370, 1991.

[11] Z. Ding, C. W. J. Granger, and R. F. Engle, "A long memory property of stock market returns and a new model," *Journal of Empirical Finance*, vol. 1, no. 1, pp. 83–106, 1993.

[12] D. Creal, S. J. Koopman, and A. Lucas, "Generalized autoregressive score models with applications," *Journal of Applied Econometrics*, vol. 28, no. 5, pp. 777–795, 2013.

[13] E. N. Nortey, K. Asare, and F. O. Mettle, "Mettle, Extreme value modelling of Ghana stock exchange index," *SpringerPlus*, vol. 4, no. 1, pp. 696–717, 2015.

[14] P. Yao, J. Wang, A. J. Yang, and J. G. Lin, "GAS volatility model based on SGT distribution and its application in risk forecasting," *Journal of Applied Sport Management*, vol. 39, no. 5, pp. 884–892, 2020.

[15] C. Rongda and J. Xu, "Forecasting volatility and correlation between oil and gold prices using a novel multivariate GAS model," *Energy Economics*, vol. 78, no. 5, pp. 379–391, 2019.

[16] E. Lazar and X. Xue, "Forecasting risk measures using intraday data in a generalized autoregressive score framework," *International Journal of Forecasting*, vol. 36, no. 3, pp. 1057–1072, 2020.

[17] B. Gnedenko, "Sur la distribution limite du terme maximum d'une serie aleatoire," *Annals of Mathematics*, vol. 44, no. 3, pp. 423–453, 1943.

[18] F. M. Longin, "The asymptotic distribution of extreme stock market returns," *Journal of Business*, vol. 69, no. 3, pp. 383–408, 1996.

[19] C. Dombry and A. Ferreira, "Maximum likelihood estimators based on the block maxima method," *Bernoulli*, vol. 25, no. 3, pp. 1690–1723, 2019.

[20] E. Balaban, T. Ozgen, and M. S. Girgin, "Distributional characteristics of interday stock returns and their asymmetric conditional volatility: firm-level evidence," *Physica A: Statistical Mechanics and Its Applications*, vol. 508, pp. 280–288, 2018.

[21] C. M. Jarque and A. K. Bera, "Efficient tests for normality, homoscedasticity and serial independence of regression residuals," *Economics Letters*, vol. 6, no. 3, pp. 255–259, 1980.

[22] F. Caeiro and M. I. Gomes, "Threshold selection in extreme value analysis," *Extreme value modeling and risk analysis: Methods and applications*, vol. 12, pp. 69–82, 2015.

[23] J. Wolfowitz, "Generalization of the theorem of glivenko-cantelli," *The Annals of Mathematical Statistics*, vol. 25, no. 1, pp. 131–138, 1954.

[24] W. H. DuMouchel, "Estimating the stable index α in order to measure tail thickness: a critique," *Annals of Statistics*, vol. 11, no. 4, pp. 1019–1031, 1983.

[25] H. Drees and M. Knežević, "Peak-over-threshold estimators for spectral tail processes: random vs. deterministic thresholds," *Extremes*, vol. 23, no. 3, pp. 465–491, 2020.

[26] B. M. Hill, "A simple general approach to inference about the tail of a distribution," *Annals of Statistics*, vol. 3, no. 5, pp. 1163–1174, 1975.

[27] S. Ghosh and S. Resnick, "A discussion on mean excess plots," *Stochastic Processes and Their Applications*, vol. 120, no. 8, pp. 1492–1517, 2010.

[28] P. Kupiec, "Techniques for verifying the accuracy of risk measurement models," *The Journal of Derivatives Winter*, vol. 3, no. 2, pp. 93–84, 1995.