Study on Operational Risk of Internet Supply Chain Finance

WANG Baosen[a],*; YANG Bin[b], MEI Panpan[b]

[a]Professor, Department of Economics, Beijing Wuzi University, Beijing, China.
[b]Postgraduate, Department of Economics, Beijing Wuzi University, Beijing, China.
*Corresponding author.

Abstract

In recent years, with the combination of "Internet + finance", traditional supply chain finance has also accelerated the online operation, and the risks it faces have gradually been exposed. This paper will use risk management theory, select risk identification and risk measurement methods, to deeply analyze the Internet supply chain financial operational risk. Enhance the cognition and emphasis of financial institutions on operational risk.

Key words: Risk measurement; POT model; Copula; VaR

INTRODUCTION

In recent years, with the deepening of the Internet and the finance, the traditional supply chain finance has also accelerated the pace of online operations, and the risks they face are gradually exposed. This paper will use the theory of financial risk management to select the appropriate online supply chain financial operation risk identification method, classify the operational risk and constructs the operational risk identification index system, and use the model to measure the operational risk in the online supply chain finance, so that it could take the financial institutions' attention to management operational risk.

1. POT MODEL

The POT (Peak over Threshold) model builds a model of sample data that exceeds a given threshold and then depicts the tail distribution of risk factors (Chen, Yang, & Huang, 2003). In the case of Internet supply chain financial operation risk, the risk events that bring huge asset losses to financial institutions basically show the characteristics of low frequency and high loss. The related losses caused by such risk events often show extreme values. POT model does not have high demand for loss historical data, so it is more suitable for measuring China's Internet supply chain financial operational risk data which is not enough for serious situation.

1.1 The Establishment of POT Model

The specific modeling process is as follows:

Let $X_1, X_2, \cdots, X_n$ be a set of random samples that are independent and identically distributed to the population X and represent the different types of loss events that are collected. Suppose the distribution function of population X is $F(x)$.

In the collected loss data, select a sufficiently large loss amount as the threshold u, $Y=X-u$ as an excess number, $F_u(y)$ can be defined as:

$$F_u(y)=P(X-u \leq y \mid X > u), y \geq 0 . \qquad (1)$$

$F_u(y)$ is the distribution function in X over the threshold u part:

Calculated by conditional probability rules available

$$F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} = \frac{F(x) - F(u)}{1 - F(u)} . \qquad (2)$$

So we can get:

$$F(x)=F_u(y)(1-F(u))+F(u) . \qquad (3)$$

Suppose y is the data distribution beyond the threshold. According to the theorem of extreme value theory, we can get $F_u(y) \approx G_{\xi,\sigma(u)}(y)$, that is Y follows the generalized Pareto distribution (GPD), as is shown in Formula (4):

$$G_{\xi,\sigma}(y) = \begin{cases} 1 - \left(1 + \frac{\xi}{\sigma}y\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \exp\left(-y/\sigma\right), & \xi = 0 \end{cases} \quad (4)$$

The shape parameter and scale parameter of the GPD function of Internet supply chain financial operation risk are represented by the Greek alphabet ξ and (Zhang & Zhou, 2015). For a given threshold u, the probability density function can be calculated by deriving Y, and then the parameters in the GPD function can be obtained by using the maximum likelihood estimation to estimate the probability density function.

According to the theorem, for any

$$\lim_{u \to x_0} \sup_{0 < y < x_0 - u} \left| F_u(y) - G_{\xi,\sigma_{(u)}}(y) \right| = 0. \quad (5)$$

From Formula (6)

$$F_u(y) = G_{\xi,\sigma(u)}(y), \quad (6)$$

we get the approximate distribution of the function:

$$F(x) = G_{\xi,\sigma(u)}(y)(1 - F(u)) + F(u). \quad (7)$$

For the data less than the threshold u, the approximate distribution $F(x)$ is usually obtained by empirical distribution. The empirical distribution function is: $F(u) = n - N_{u/n}$, where is the number of samples, which is beyond the threshold u of N_u and is substituted into Equation (4)

$$F(x) = \begin{cases} 1 - \frac{N_u}{N}\left(1 + \frac{\xi}{\sigma}(x - u)\right)^{-1/\xi}, & \xi \neq 0 \\ 1 - \frac{N_u}{N}\exp\left(-(x-u)/\sigma\right), & \xi = 0 \end{cases} \quad (8)$$

Based on the above discussion, we can get the distribution function of the four types of operational risk of internet supply chain finance beyond u. Finally, based on the VaR theory, we further calculate the VaR of the four types of operational risk and obtain the corresponding VaR and ES under the given confidence P:

$$\text{VaR} = \begin{cases} u + \frac{\sigma}{\xi}\left\{\left[\frac{N}{N_n}(1-p)\right]^{-\xi} - 1\right\}, & \xi \neq 0 \\ u - \sigma\ln\left[\frac{N}{N_n}(1-p)\right], & \xi = 0 \end{cases}, \quad (9)$$

$$ES = \frac{\text{VaR}}{1 - \xi} + \frac{\sigma - \xi u}{1 - \xi}. \quad (10)$$

According to the above process, the four types of operational risk loss data of Internet supply chain finance can be more fully estimated the tail distribution that exceeds u.

1.2 Determination of Threshold in POT Model

In this paper, we use the Hill graph method and the graph of the excess mean function to determine the threshold u,

and then use the maximum likelihood estimation method to calculate two types of parameters in the POT model, and then use the goodness of fit to select the optimal u.

The method of determining the threshold u is as follows:

(a) Hill chart method

The overall X is expressed in the form of order statistics, we can get $X_1 > X_2 > \cdots > X_n$. Hill statistics, as is shown in Formula (11):

$$\xi_{k,n} = \frac{1}{k}\sum_{i=1}^{k} \ln\left[\frac{X(i)}{X(k)}\right], \quad 1 \leq k \leq n. \quad (11)$$

The Hill plots are scatter grams made up of points $\left(k, 1/\xi_{k,n}\right)$. Usually we use the X corresponding to the x-coordinate k of the start of the more stable part of the Hill plot as the threshold u.

(b) Excess map method

Excess mean function is defined as $e(u) = E(X-u \mid X>u)$, and we can get the sample excess function:

$$e(u) = \sum_{i=1}^{n}(x_i - u)/N_u. \quad (12)$$

Excess mean plots are scatter grams made up of points $(u, e(u))$. is the amount of samples in the collected data that exceeds the threshold u. If a sufficiently large threshold u is chosen, when $\geq u$, it approximates to a linear function. The parameter $\xi > 0$ indicates that the data presents a thick tail distribution and obeys the generalized Pareto distribution. But when $x \leq u$ and the parameter $\xi < 0$ indicate that the collected data do not have a thick tail feature.

(c) The optimal threshold is chosen by the χ^2 goodness of fit

Based on the Pickands-Balkama-de Haan theorem, when the threshold u goes to infinity, the conditional distribution function that exceeds the threshold converges to the GPD distribution. The threshold near the conditional distribution and the GPD distribution of the samples that exceed the threshold is the best threshold.

For different thresholds, a new quantity combination is built up for the loss data beyond the threshold and then to calculate the χ^2 statistics of different thresholds u,

$$\chi^2 = \sum_{i=1}^{N}(N_i - np_i)^2/np_i.$$

The calculated χ^2 value can be used to select the threshold. The smaller the value, the higher the confidence. That is, the better the fitting result is, the better the selected threshold is, and vice versa.

1.3 POT Model Parameter Estimation

After the four kinds of operational risk thresholds are calculated, two parameters in the GPD distribution are estimated and the maximum likelihood estimation method is used. Specifically, the probability density function of GPD can be obtained by Formula (4):

2

$$g_{\xi,\sigma}(y) = \begin{cases} \frac{1}{\sigma}\left(1+\frac{\xi}{\sigma}y\right)^{-\left(1+1/\xi\right)}, & \xi \neq 0 \\ \frac{1}{\sigma}\exp(-y/\sigma), & \xi = 0 \end{cases} \quad (13)$$

Thus, the log-likelihood function in Equation (14) is obtained:

$$L(\xi, \sigma; y) = \begin{cases} -n\ln\sigma - \left(1+\frac{1}{\xi}\right)\sum_{i=1}^{n}\ln\left(1+\frac{\xi}{\sigma}y_i\right), & \xi \neq 0 \\ -n\ln\sigma - \frac{1}{\sigma}\sum_{i=1}^{n}y_i, & \xi = 0 \end{cases} \quad (14)$$

For the above logarithmic likelihood function, let

$$\begin{cases} \partial l/\partial\xi = 0 \\ \partial l/\partial\sigma = 0 \end{cases},$$

we can deduce a likelihood equation group about ξ and σ and then solve the equation group to obtain the estimated value of the parameter ξ and σ.

2. THE INTEGRATED MEASUREMENT OF INTERNET SUPPLY CHAIN FINANCE OPERATIONAL RISK BY INTRODUCING COPULA FUNCTION

The Copula function describes the correlation between variables and is essentially the method of calculating the joint distribution function by using the edge distribution of random vectors to make it more in line with the interrelated characteristics of different types of operational risk in the supply chain finance of the Internet. Therefore, the introduction of Copula function can provide a more accurate measurement of the operational risk of Internet supply chain (Liu, 2014).

2.1 The Introduction of Copula Function

Let *H* be a joint distribution of n dimensions, the edge distribution is *F*1, *F*2, ..., *Fn* respectively. Then for all *X*, there is an n-dimensional Copula function C such that:

$$H(x_1, x_2, \cdots, x_n) = C(F(x_1), F(x_2), \cdots, F(x_n)). \quad (15)$$

On the other hand, if *C* is an n-dimensional Copula function and *F*1, *F*2, ..., *Fn* are distribution functions, then there exists an *n*-dimensional joint distribution function *H* such that Formula (15) holds and the edge distribution functions are *F*1, *F*2, ..., *Fn*.

Formula (15) about the joint partial derivative, we get:

$$h(x_1, x_2, \cdots, x_n) = c(F(x_1), F(x_2), \cdots, F(x_n))\prod_{i=1}^{n}f_i(x_i), \quad (16)$$

where

$$c(F(x_1), F(x_2), \cdots, F(x_n)) = \frac{\partial C(F(x_1), F(x_2), \cdots, F(x_n))}{\partial F(x_1)\partial F(x_2)\cdots\partial F(x_n)}.$$

Common Copula functions fall into two main categories: Elliptical Copula and Archimedean Copula.

Ellipse Copula function can well depict the correlation between variables, but because of its symmetry characteristics, it can not describe the asymmetric relationship of Internet supply chain financial operation risk well, so it is not suitable for measuring China's Internet supply chain finance integrated operational risk.

(a) Clayton Copula function

Take the two-dimensional Clayton Copula distribution function as an example:

$$C(u,v;\theta) = Max((u^{-\theta} + v^{-\theta} - 1)^{\frac{-1}{\theta}}, 0). \quad (17)$$

Where θ is the dependent parameter, whose range is $[-1, \infty] \cup \{0\}$.

(b) Gumbel Copula function

Take the two-dimensional Gumbel Copula distribution function as an example:

$$C(u,v;\theta) = \exp\{-[(-\log(u))^{\theta} + (-\log(v))^{\theta}]^{\frac{1}{\theta}}\}. \quad (18)$$

Where θ is the dependent parameter, whose range is $[1, \infty)$.

(c) Frank Copula function

Take the two-dimensional Frank Copula distribution function as an example:

$$C(u,v;\theta) = -\theta^{-1}\log(1 - \frac{(1-e^{-\theta u})(1-e^{-\theta v})}{(1-e^{-\theta})}). \quad (19)$$

Where θ is the dependent parameter, whose range is $\theta \neq 0$.

Because of the financial institutions in the Internet supply chain finance operational risk, various risk loss data tail distributions show a certain relevance, and the Copula Copula Gumbel Copula function and Clayton Copula function not only can better depict all kinds of operations Thick-tail risk and asymmetry, and the calculation is more concise. Therefore, this section uses the Gumbel Copula function and the Clayton Copula function to study the correlation between different operational risks in internet supply chain finance.

2.2 Copula Function Parameter Estimation

Since the edge distribution of the four types of operational risk of Internet supply chain finance is determined by using POT, the two-stage method is adopted to estimate the parameters of Copula function, thus estimating the value of the parameter and .

2.3 Optimal Copula Function Selection

The appropriate Copula function can better depict the correlation between the four types of operational risk. Since Gumbel Copula function and Clayton Copula function in Archimedean Copula are chosen, according to the experience of Nelsen, Q-Q graph method can be used to judge the optimal Copula function when selecting Archimedean Copula.

Observing the distribution of Q-Q diagram, if it is approximately a straight line, it means that the fitting

effect of the function is more realistic, so the optimal Copula function can be selected according to this standard. Therefore, the Copula function, which best shows the correlation between the four types of operation risk of Internet supply chain finance, can be selected through the Q-Q diagram. Thus, the Copula-POT comprehensive measurement model of Internet supply chain financial operation risk is constructed.

3. EMPIRICAL RESEARCH

3.1 Data Sources and Statistical Analysis

CITIC Bank supply chain financial services in the industry-leading, and is a listed company, the data is easier to obtain. After a search and collation, a total of 65 event samples of loss caused by operational risk of supply chain finance business of China CITIC Bank from 2005 to 2015 were obtained. Based on the FTA method of Internet supply chain financial operational risk, the 65 data are divided into four types of loss events: online financing, e-commerce, logistics transaction and online payment transaction risk, corresponding to the loss event types I, II, III and IV respectively, Specific statistics in Table 1.

Table 1
CITIC Bank Supply Chain Financial Operational Risk Loss Statistics   (Unit: ten thousand yuan)

	Number of samples	Mean	Maximum	Minimum	Standard deviation
Overall	65	41,353	310,000	5	49,093
Loss event type I	28	63,972	121,130	10	22,456
Loss event type II	11	5,618	11,145	5	3,780
Loss event type III	9	2,491	7,000	5	2,120
Loss event type IV	17	47,794	310,000	5	78,170

Before applying the model, the Q-Q diagram is usually used to test the thick-tail characteristics of the sample data to determine the applicability of the model. When the Q-Q diagram is curved upward, it indicates that the sample data has a thick tail characteristic. When the Q-Q diagram shows the downward bending state, the sample data has a thin tail characteristic. Through SPSS

analysis, the four types of loss event type QQ chart is shown in Figure 1, and the four QQ charts are bent upward trend, which can determine the four types of Internet supply chain financial operational risk loss data with thick tail distribution Feature, you can use the POT model to measure the Internet supply chain financial operational risk.
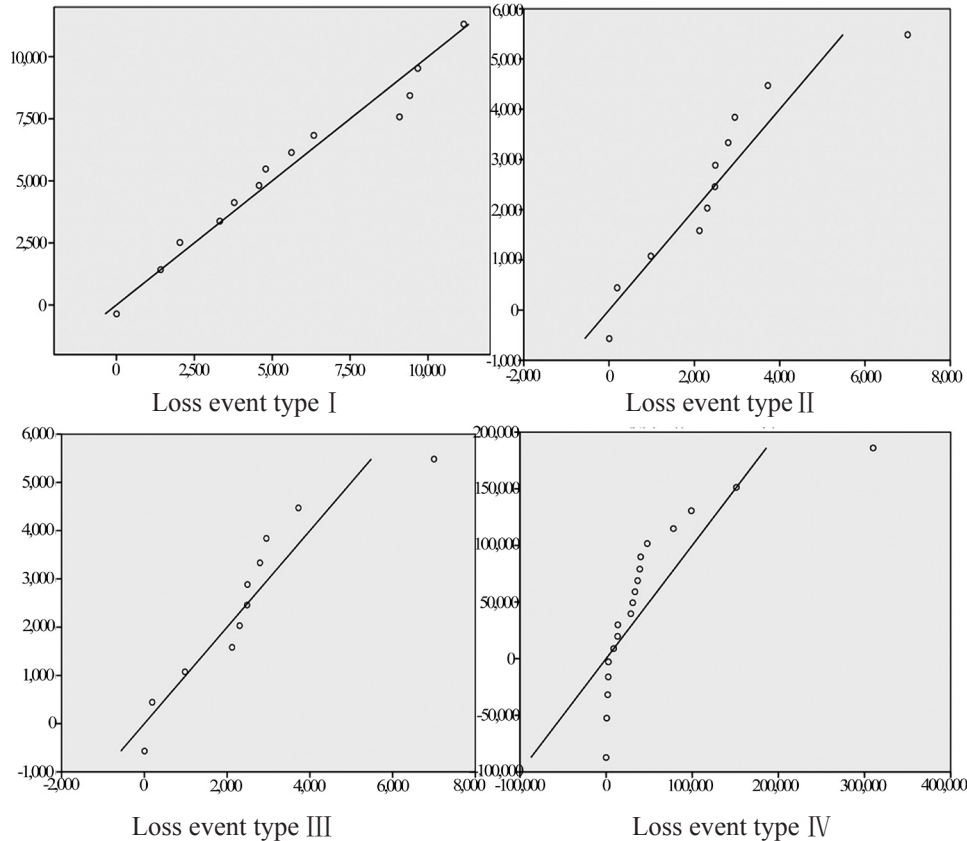

Loss event type I

Loss event type II

Loss event type III

Loss event type IV

Figure 1
Q-Q Diagram of Operational Risk Loss of Internet Supply Chain Finance of CITIC Bank

3.2 Application of POT Model in Internet Supply Chain Financial Operation Risk Measurement

Under the 95% confidence interval, Figure 2 and Figure 3 correspond to Hill and excess mean values for the four types of operational risk loss event thresholds respectively. Figures (a), (b), and (c), (d) Corresponding to Internet Supply Chain Financial Operational Risk Classification Loss Type Loss Event Type I, II, III, IV.

Observe the graph (a) in Figure 2. When the excess number is 3, the Hill graph gradually assumes a steady state, and the X value corresponding to the abscissa k of the point is the threshold value. In the same way, there may be three thresholds corresponding to the other three types of operational risk loss event types, as shown in the other figures of Figure 2.
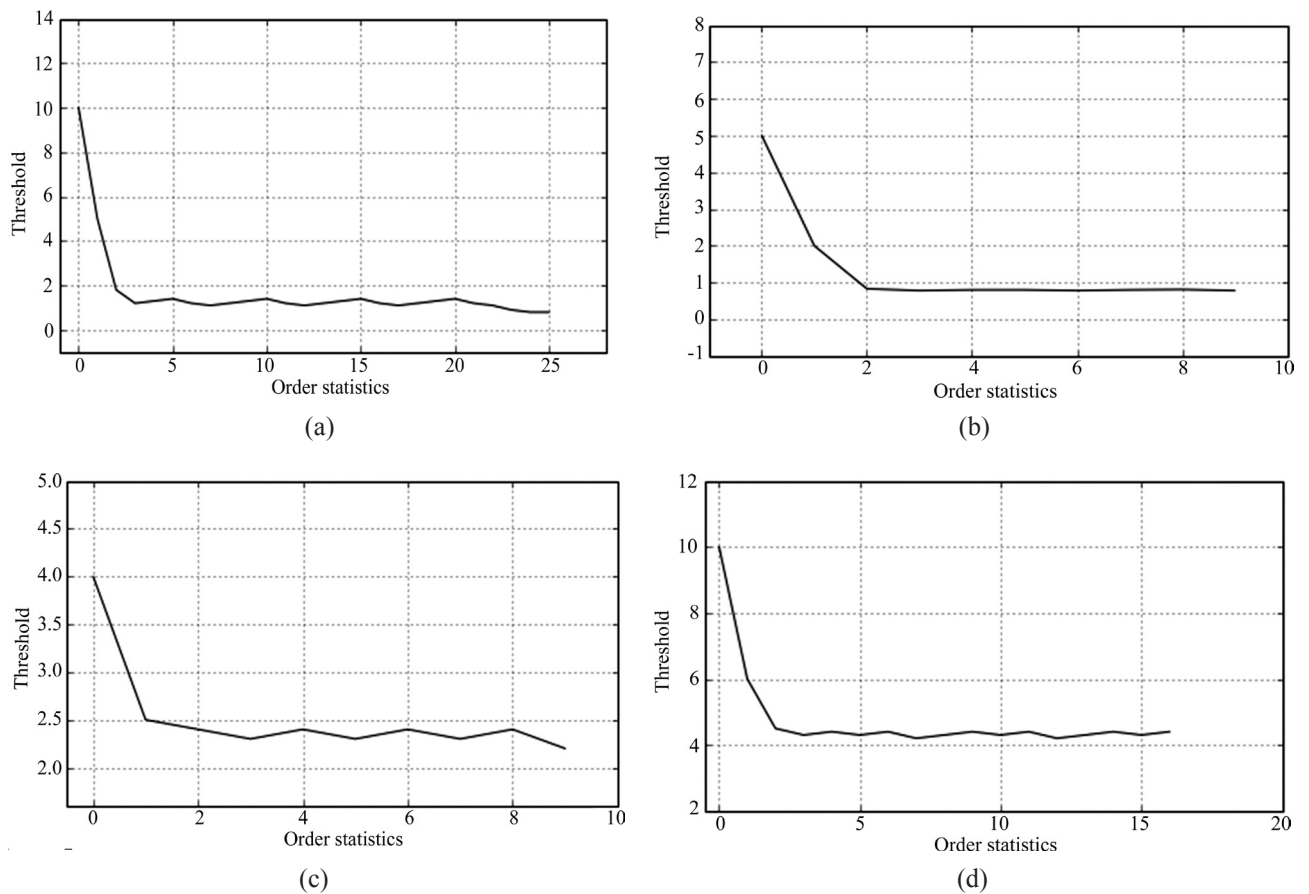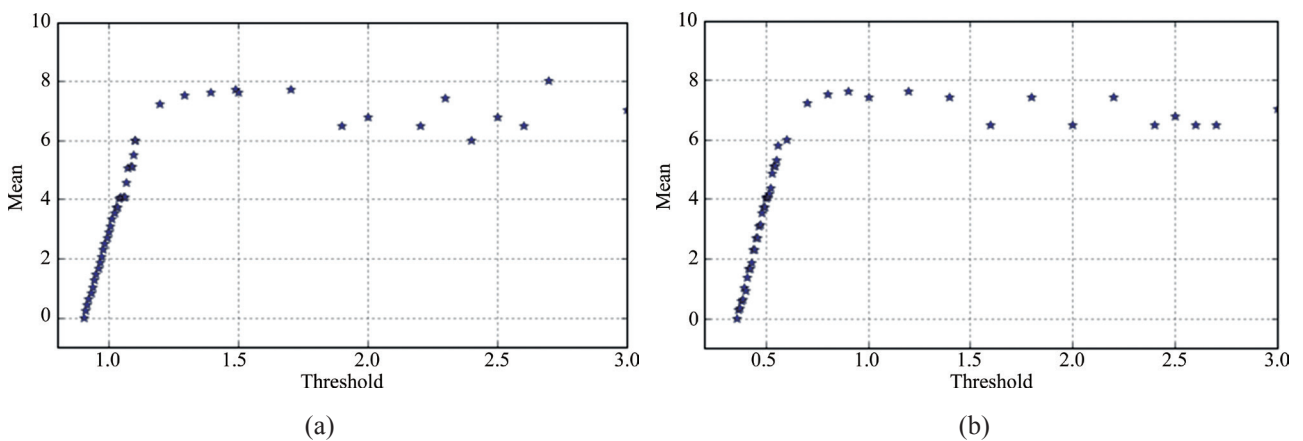

(a)

(b)

(c)

(d)

Figure 2
Hill Diagram of Four Types of Loss of Operational Risk
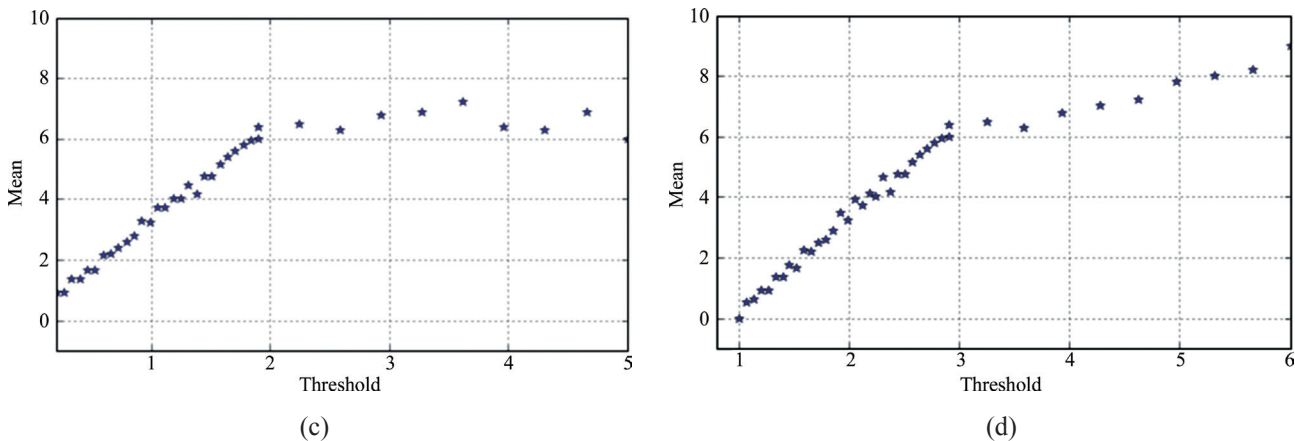

(a)

(b)

(c)

(d)

Figure 3
Excess Averages of Four Types of Loss of Operational Risk

Similarly, in Figure 3, the excess mean value maps corresponding to the four types of operational risk loss types are respectively shown, in which the abscissas of each graph correspond to the amount of loss is times of the abscissa scale. The estimated curve gradually shows the starting point of the linear trend, and the value corresponding to the abscissa of the point is multiplied by the threshold.

After the Hill graph and the excess mean value graph respectively, the thresholds of various types of operational risk losses in the Internet supply chain finance are shown in Table 2.

Table 2
Threshold Selection Table for Four Types of Operational Risk

Different types	Hill diagram method	Excess mean plot method
Loss event type I	1,800	1,124
Loss event type II	813	756
Loss event type III	250	198
Loss event type IV	450	297

According to the optimal threshold selection method in the previous chapter, the goodness-of-fit test is performed on each threshold obtained in Table 2. The smaller the corresponding threshold is, the better the threshold is obtained. For the parameter and in the POT model, the maximum likelihood estimation method is used to estimate $F(x)$. In this way, the parameter estimation value corresponding to each threshold is obtained. Table 3 shows the optimal threshold and parameter estimation results of the four types of loss types.

Table 3
Determination of the Optimal Threshold of the Four Types of Operational Risk

Different types	Threshold u	The number Nu beyond the threshold u	Parameter estimation	Parameter estimation
Loss event type I	1,800	16	0.8489	6,037
Loss event type II	756	6	0.8571	1,050
Loss event type III	250	4	0.5697	8,507
Loss event type IV	450	9	1.2589	4,839

The data in Table 3 are substituted into the formula for calculating VaR (9), and the VaR of CITIC Bank under various confidence levels is calculated one by one according to the results shown in Table 4.

Table 4
VaR Value of Four Operational Risks (Unit: ten thousand yuan)

Different types	95%	99%
Loss event type I	25,916.84	50,934.75
Loss event type II	4,774.59	9,029.25
Loss event type III	20,247.04	37,160.40
Loss event type IV	27,935.28	71,580.88

3.3 Application of Copula Operational Risk Comprehensive Measurement Model After Joining Copula

After the above analysis of selecting the optimal Copula, this section will adopt the two Copula functions in the Archimedean Copula function to measure the integration risk of Internet supply chain financial operations. The function forms are: Clayton Copula function and Gumbel Copula function.

3.3.1 Copula Function Parameter Estimation

After the above study on Copula function parameter estimation method, this section uses two-stage method to estimate the parameters of Clayton Copula function and Gumbel Copula function.

The parameters estimation of Clayton Copula function and Gumbel Copula function are shown in Table 5. As can be seen from the table, both types of Copula function parameters are greater than 1, so there is a significant correlation between various operational risks in Internet supply chain finance.

Table 5
Copula Function Parameter Estimation

Copula function	Gumbel copula	Clayton copula
Estimate parameters	1.7829	2.4918

3.3.2 Optimal Copula Function Selection

Using Q-Q graph method to determine the optimal Copula function. As can be seen from Figure 4, Gumbel Copula is most similar to the Q-Q graph and the straight line. The Q-Q graph of Clayton Copula has a large deviation from the straight line, indicating that the Gumbel Copula function has the best fit to the joint loss distribution effect.
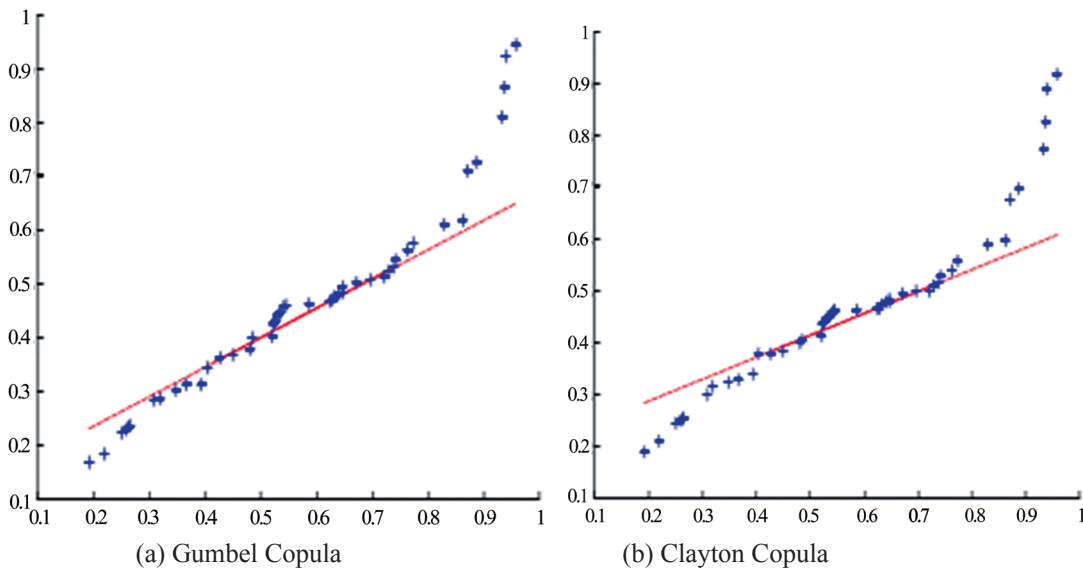

(a) Gumbel Copula   (b) Clayton Copula

Figure 4
Q-Q plot of Copula Function

Based on the above analysis, according to Formulas (8) and (19), we finally get the joint distribution function of four types of operational risk of CITIC Bank in supply chain financial business:

$$\text{F}(x_1, x_2, x_3, x_4) = C_\theta(u_1, u_2, u_3, u_4) = C_{1.7829}\big(F_1(x_1), F_2(x_2), F_3(x_3), F_4(x_4)\big).$$

Where
$$C_\theta(u_1, u_2, u_3, u_4) = \exp\left\{-\big[\textstyle\sum_{i=1}^{4}(-\log(u_i))^\theta\big]^{\frac{1}{\theta}}\right\}, \tag{20}$$

$$F_1(x_1) = 1 - 0.57[1 + 0.00014(x - 1800)]^{-\frac{1}{0.8489}}, \tag{21}$$

$$F_2(x_2) = 1 - 0.55[1 + 0.00082(x - 756)]^{-1/0.8571}, \tag{22}$$

$$F_3(x_3) = 1 - 0.44[1 + 0.000067(x - 250)]^{-1/0.5697}, \tag{23}$$

$$F_4(x_4) = 1 - 0.53[1 + 0.00026(x - 450)]^{-1/1.2589}. \tag{24}$$

Cupula function, which is the joint distribution of the marginal distribution of the four types of operational risk in Internet supply chain finance, where θ is the parameter value of Gumbel Copula function, which is 1.7829, and is the distribution function of the *i*-th class of operational risk. Therefore, by introducing the Gumbel Copula function empirically, the correlation between the four types of operational risk is considered, so that the POT model can measure the total capital of operational risk regulation more accurately.

3.3.3 Introducing Copula Function and Monto Carlo Simulation to Calculate VaR

Since we have obtained the marginal distribution parameters of various types of operational risk and the Gumbel Copula function in the internet supply chain finance and the correlation between the four types of operational risk losses in supply chain finance, Monto Carlo is used to measure the comprehensive operational risk capital.

To sum up the above four types of operational risk losses in the Internet supply chain finance, the integrated VaR of operational risk is obtained, which is compared with that of the above without considering the correlation. The comparison results are shown in Table 6 .

Table 6
VaR Without Considering the Correlation and Considering the Correlation   (Unit: ten thousand yuan)

	Regardless of the relevance of the VaR	Consider the VaR under the correlation	Percentage reduction
99%	168,705.28	15,2948.2	9.34%
95%	78,873.74	70,134.53	11.08%

It can be seen that the introduction of Copula function can avoid the decentralization effect of operational risk, reduce the regulatory risk of commercial banks effectively, improve Pareto for commercial banks and make full use of capital. And, when the level of confidence reaches a high level, the probability of an Internet supply chain financial operational risk event has dropped to a very low level. China's commercial banks should maintain a higher level of confidence in order to better reduce the probability of occurrence of operational risks in the supply chain of Internet.

REFERENCES

Chen, X. H., & Yang, H. Y., & Huang, X. Y. (2003). The POT model for operational risk in banks. *Management Science in China,* (1), 49-52.

Liu, S. W. (2014). *Measurement of operational risk in commercial bank based on copula-POTT method*. Harbin Institute of Technology.

Zhang, H. X., & Zhou, Y. Y. (2015). Loss distribution of flood catastrophes based on POT-GPD model and comparison of parameter estimation methods. *Journal of Henan Science and Technology,* (2), 151-153.