A Study on Operational Risk and Credit Portfolio Risk Estimation Using Data Analytics

Rongda Chen

School of Finance, Zhejiang University of Finance and Economics, Hangzhou 310018, China and Center for Research of Regulation and Policy of Zhejiang Province, Hangzhou 310018, China, e-mail: rongdachen@163.com

Ze Wang

School of Finance, Zhejiang University of Finance and Economics, Hangzhou 310018, China, e-mail: ze1188@hotmail.com

Liu Yang[†]

Business School, University of International Business and Economics, Beijing 100029, China, e-mail: yangliu@uibe.edu.cn

Chi To Ng and T.C.E. Cheng

Department of Logistics and Maritime Studies, Logistics Research Centre, The Hong Kong Polytechnic University, Hong Kong SAR, China, e-mail: daniel.ng@polyu.edu.hk, Edwin.Cheng@polyu.edu.hk

[†]Corresponding author.

A Study on Operational Risk and Credit Portfolio Risk Estimation Using Data Analytics

Abstract

In this paper we consider operational risk and use data analytics to estimate the credit portfolio risk. Specifically, we consider situations in which managers need to make the optimal operational decision on total provision for risk to hedge against the potential risk in the entire supply chain. We build a new structural credit model integrated with data analytics to analyze the joint default risk of credit portfolio. Our model enables the decision maker to better assess the risk of a supply chain, so that they could determine the optimal operational decisions with total provision for risk, and react in a timely manner to economic and environmental changes. We propose an efficient simulation method to estimate the default probability of the credit portfolio with the risk factors having the multivariate *t*-copula. Moreover, we develop a three-step importance sampling (IS) method for the *t*-copula credit portfolio risk measurement model to achieve an accurate estimation of the tail probability of the credit portfolio loss distribution. We apply the Levenberg-Marquardt algorithm to estimate the mean-shift vector of the systematic risk factors after the probability measure change. Besides, we empirically examine the changes in the credit portfolio risks of 60 listed Chinese firms in different industries using our proposed method. The results show that our model can help the decision maker make the optimal operational decisions with total provision for risk, which hedges against the potential risk in the entire supply chain.

Keywords: operational risk management; data analytics; decision making; simulation; credit portfolio risk

1 Introduction

Firms' ability in risk management significantly affects their operational decisions, especially when facing rapid economic and environmental changes, and technology advancement. Risk management affects firms' operational decisions from different perspectives. A good understanding of risk helps firms avoid significant losses and make the optimal operational decisions. On the contrary, poor estimation of risk leads to inferior operational decisions and serious losses to firms. Given that each firm's operational decisions have risk, estimating the risk and preparing reasonable risk provision are essential. As for industries, there must be connections among different firms, no matter they are competitors or cooperators. Once one firm faces a crisis, the crisis will be spread to its upstream or downstream firms. It is necessary to have the right risk provision for helping firms to overcome the crisis. For a government or a central bank, it is also essential for it to prepare enough risk provision for the situation when the firms of all industries default. Therefore, for firms or leaders of the economy, they all need to make a decision to prepare the provision of adequate money when a crisis arises. For instance, Ofo, one of the biggest bicycle-sharing firms in China, has ceased its operations in many nations, including Germany, Australia, Israel, Austria, and India, and made massive reductions in operations in other countries due to its worsening financial status. The firm was also sued by its major supplier, a bicycle manufacturing firm, for unpaid bill (National Business Daily, 2018). Therefore, firms need to manage financial risk well in order to maintain correct operational decisions. In many industries, the decision makers need to make operational decisions with appropriate total provision for risk to hedge again risk in the entire supply chain.

Therefore, financial risk management and operational risk management are closely related to each other. For example, when supply disruption occurs due to some disasters, like earthquake or flood, firms have to use their provision to hedge against the caused damage. When the market demand suddenly declines due to some unpredictable reasons, like a flu epidemic, firms also need enough provision to cope with the crisis. On the other hand, too much provision saved will affect firms' investment efficiency, which will in turn significantly affect firms' performance. Therefore, how to calculate the right amount of total provision for risk is essential to the management of operational risk in the entire supply chain. Since total provision for risk has an inherent relationship with credit portfolio risk, credit portfolio risk estimation is key to calculating the right amount of total provision for risk in the entire supply chain. Thus, a comprehensive and sound understanding of the credit portfolio risk is a prerequisite for effective operational risk management.

To achieve the optimal operational decisions with total provision for risk, two issues are critical, namely data collection and risk estimation. Considering the huge amount of data available with the rapid advances of information technology, it has become important and feasible to examine risk by using data analytics (Koyuncugil and Ozgulbas, 2012; Chen et al., 2015; Choi et al., 2017; Kou et al., 2019; Sun et al., 2019; Wang and Wu, 2020). Credit default events are rare and thus the probability of large losses of credit portfolio is small. Calculating the probability of default of rare events and the corresponding losses are crucial to risk management. Accurately estimating the tail probability distribution in the credit portfolio risk distribution enables firms to fully prepare the credit portfolio risk loss reserve in a timely manner. For banks, when the asset portfolio is a loan portfolio, then banks can simply retain reasonable funds to meet the provision coverage ratio without excessively placing idle funds, thereby maximizing the bank's profitability while satisfying its risk preparation. The corresponding research on estimating firms' credit portfolio risk has been emphasized in both academia and industries (Hailemariam et al., 2012; Boudreault, 2015; Hsieh et al., 2018; Osmundsen, 2018; Chen et al., 2018; Bülbül et al., 2019).

In this paper we develop a data analytics technique to generate some useful information about credit portfolio risk by exploiting data from different sources. It is worth noting that our model is a general model, which can be applied to any industries. In this paper we apply it to three industries, namely the real estate industry, retail industry, and finance and insurance industry. They are classified as industries in the Shanghai Stock Exchange in China. First, in the computational experiments, using the classical structural credit model, we integrate the six dimensions of data embracing firm assets, firm volatility, firm debts, firm leverage ratios, return on assets, and interest rate level to calculate the default probability of an individual firm. By doing so, we are able to understand more about the firms' credit risks. Second, we integrate the multivariate *t*-copula into the classical structural credit model to calculate the joint default probability of the credit portfolio. Simultaneously, we integrate multiple dimensions of data to obtain the joint default probability of the credit portfolio. Thus, the joint default probability contains information from various sources. Third, we further incorporate the achieved Expected Shortfall (ES) value to determine the optimal operational decision on total provision for risk. Thus, the total provision is obtained by integrating the multivariate *t*-copula into the classical structural credit model, and using the proposed three-step importance sampling (IS) model. Our integrated model can be widely applied in different situations, and the corresponding results enable firms to better assess the financial risks of themselves and their suppliers/customers. As such, the firms could make proper operational decisions in a timely manner, especially when facing rapid economic and environmental changes, and technology advancement. The above explains how we apply the proposed data analytics technique in three aspects, i.e., synergizing the classical structural credit model, integrating the multivariate *t*-copula with the classical structural credit model, and using the proposed three-step importance sampling (IS) model, to estimate credit portfolio risk in the framework of operational risk management.

Further, we extend the two-step importance sampling (IS) technique to a three-step IS technique for the *t*-copula credit portfolio risk measurement model in order to further reduce the variance. Note that choosing the mean-shift vector of the systematic risk factors after the probability measure change undermines the effectiveness of the IS technique when solving the optimization problem. The Gauss-Newton method is usually used to solve the optimization problem. But the Gauss-Newton method has the shortcoming of slow convergence, which needs improvement. To address this issue, we apply the Levenberg-Marquardt algorithm, a nonlinear optimization technique, to improve the solution. The algorithm combines the merits of the Gauss-Newton algorithm and the gradient approach. This is the main difference between our model and the previously proposed models in the literature, such as Kang and Shahabuddin (2005), Kostadinov (2006), Bassamboo et al. (2008), Chan and Kroese (2010), and Reitan and Aas (2010), in the context of determining the optimal mean-shift vector.

In addition, Value-at-Risk (VaR) lacks the property of subadditivity when the risk factors have heavy-tailed distributions. As VaR is sensitive to changes in the significance level, it is not a coherent risk measure. This leads to failures in achieving the numerical stability needed for bank management decisions (Frey and McNeil, 2002; Kalkbrener et al., 2004; Broda, 2012). In view of these drawbacks of VaR, we apply expected shortfall (ES), i.e., the expected excess loss given that there are portfolio losses, to measure credit portfolio risk. Since VaR and ES are generally based on the same estimated loss distribution, ES is closely related to VaR. Meanwhile, the Basel III Accord (the Third Basel Accord) suggests that ES should be applied to complement VaR for estimating the probabilities of portfolio losses. Therefore, we introduce the two risk measures VaR and ES to measure credit portfolio risk in the multivariate *t*-copula framework and compare their functionality.

Finally, differing from most of the studies on estimating credit portfolio risk that are focused on numerical computation, we empirically examine the data of 60 listed Chinese firms in different industries, taking into consideration an inherent relationship between dependent default and asset returns in the classical structural model. Specifically, we use data from 25 real estate firms, 20 retail firms, and 15 finance and insurance firms listed in the Shanghai Stock Exchange in China covering the period from 4 January 2012 to 7 June 2013. During this period, the real estate industry in China faced a new round of regulation whereby five new national real estate regulation policies were promulgated on 20 February 2013, stipulating home buying restrictions and loan limitations. Applying the proposed model to compute the tail probabilities of the credit portfolio loss distributions of these 60 listed firms and their corresponding VaR and ES values, we obtain large variance reductions. We also check the changes in credit portfolio risk in different industries corresponding to the new national real estate regulation policies promulgated in China in the same period to see how well the analytical results align with the historical events that transpired.

We organize the rest of the paper as follows: In Section 2, we review the related literature and identify the research gap. In Section 3, we apply the classical structural credit model to calculate the default probability of each firm. In Section 4, we discuss ways to derive the default probability of the credit portfolio. In Section 5, we develop a three-step importance sampling for the credit portfolio measurement model based on the *t*-copula. In Section 6, we explore the use of the proposed model to assist managers to hedge against risk in the entire supply chain. In Section 7, we conduct an empirical study using real data to assess the performance of the proposed model. In Section 8, we conclude the paper and suggest future research topics.

2 Literature Review

Our study is closely related to four streams in the literature. The first stream is about the interface between operational management and financial risk management. The second stream relates to using the data analytics to hedge against the operational risk. The third stream is about credit portfolio risk measurement. The fourth stream is about how to select a suitable copula to depict the dependent default in practice. We review the four streams to identify the research gap and position our paper in this section.

We start with the first stream on the interface between operational management and risk management. Based on financial portfolio theory, Kumar and Park (2019) proposed an integrated approach to study a variety of relationships between supply chain risk, risk management, and supply chain value. Brusset and Bertrand (2018) proposed an approach using weather index-based financial instruments to enable the risk taker to take account of weather risk and reduce sales volatility. Kouvelis et al. (2018) integrated financial hedging decision and inventory replenishment decision to maximize the mean-variance of terminal wealth over a finite horizon. Ma et al. (2019) considered the impact of risk averse on a firm's advance selling decisions and identified the conditions under which the firm prefers advance selling. Azadegan et al. (2020) illustrated a strong link between business continuity programmes (which are important in response to and making recovery from supply chain disruptions) and financial performance using fuzzy set qualitative comparative analysis. These studies demonstrate the close relationship between financial risk management and operational risk management. Particularly, firms have to consider their financial status when making operational decisions to hedge against risk. Our study provides a method to better estimate firms' financial risk in multiple dimensions, based on which the firms are able to calculate the appropriate provision for operational risk.

The second stream of the related literature concerns the adoption of data analytics to hedge against operational risk. With the rapid development of data analytics, the technique is widely applied to generate useful information by exploiting data from different sources. Academia and industries are making great efforts to mitigate operational risk by using data analytics (Leveling et al., 2014; Choi et al., 2017; Goel et al., 2017). Sun et al. (2020) study the application of data analytics to mitigate operational risk in the airline industry. Wang and Yao (2019) use data analytics to investigate how to jointly optimize the capacity decision and hedging decision. Ivanov et al. (2018) discuss the relationship between data analytics and supply chain disruption risk management. These studies emphasize the feasibility and importance of applying data analytics in operational risk management. They also stress the importance of assessing risk appropriately when applying data analytics. Our paper contributes to this research stream by proposing a new structural credit model to better assess the joint default risk of a credit portfolio using data analytics. Our proposed model could help managers gain a comprehensive understanding of risk and make a better hedge against operational risk.

The third steam is about credit portfolio risk measurement. As the credit risk of asset transferring and trading in market increases and bank regulation tightens, businesses are urged to develop new methods to measure and manage credit risk over the past few years. Plentiful studies have investigated credit portfolio risk (e.g., McNamara, 1998; Smith et al., 1996; Hu, 2016). Many popular credit portfolio risk measurement models, such as Credit-Metrics of J.P. Morgan, KMV Portfolio Manager, CreditRisk+ of Credit Suisse First Boston, and McKinsey's Credit Portfolio View, rely on Monte Carlo simulation to calculate the tail probability of the credit portfolio loss distribution or its Value-at-Risk (VaR) at a given confidence level over a fixed time horizon. However, as credit defaults by firms are rare and the threshold value of default is large, the tail probability of the credit portfolio loss distribution is small. Consequently, to conduct standard Monte Carlo simulation requires significantly large sampling, which is computationally demanding and inefficient, especially for credit portfolios involving many trading parties.

To improve the efficiency of simulation, Merino and Nyfeler (2002) proposed a method combining the fast Fourier transforms and Monte Carlo simulation to estimate the tail probability of the credit portfolio loss distribution in the frame of conditional independence. Grundke (2007) extended the well-known credit portfolio model CreditMetrics and applied an efficient Fourier transforms method to calculate credit risk. Glasserman and Li (2005) provided a two-step importance sampling (IS) technique for the extensively used multivariate normal copula model of credit portfolio risk to overcome the difficulty of calculating the small tail probability of the credit portfolio loss distribution. To increase the frequencies of large losses in the two-step IS technique, Glasserman et al. (2008) developed the Knapsack problem and the corresponding nonlinear dynamic programming solution algorithm to find the mean vector of the systematic risk factors. Grundke (2009) developed a two-step IS technique for the credit portfolio risk measurement model based on the Fourier transforms.

However, a typical shortcoming of the above research is that the systematic risk factors are described and modelled as a multivariate normal. Note that when the coefficient of the linear correlation between two random variables is less than one, the tail dependence coefficient of the normal copula is zero. This causes the default indicator function being asymptotically independent, conditional upon the marginal default probability being small. However, this is contrary to dependent default in practice. In this paper we develop a new three-step IS technique for the *t*-copula credit portfolio risk measurement model for further variance reduction so that we can obtain the tail probability of the credit portfolio loss distribution or ES value of the credit portfolio quickly and accurately, and then we can calculate the total provision for risk based on the obtained ES value.

Our study is also closely related to the selection of a suitable copula to depict the dependent default in practice. Moreover, many empirical studies have found that the distributions of asset returns often have heavy tails and high kurtosis (see, e.g., Huang and Kou, 2014; Aldrich et al., 2016; Ankudinov et al., 2017; D'Amico and Petroni, 2018). In addition, asset returns have an intrinsic relationship with dependent default in the classical structural credit model. This indicates that the traditional normal copula assumption is not entirely valid in view of the generic features of practical data. In other words, the normal copula assumption may not be able to model the dependence between the pertinent financial variables in a realistic and satisfactory manner.

The copula can well depict more default dependence by the correlations between the latent variables. However, selecting a suitable copula is difficult for cases with small samples. Because there are nonlinear correlations among the default data, depicting the default dependence between financial variables is of great importance (Rosenberg and Schuermann, 2006; Kole et al., 2007; Biglova et al., 2009).

Besides, the multivariate normal copula is unable to fully depict the dependence between financial variables. However, the *t*-copula supports extremal dependence between financial variables, given that a useful interpretation of extremal dependence is that it obeys a multivariate *t*-distribution consisting of the square root of a scaled chi-squared random variable and a multivariate normal (Kang and Shahabuddin, 2005; Bassamboo et al., 2008; Reitan and Aas, 2010; Chan and Kroese, 2010).

Some research found that fitting the Student's *t*-distribution to risk factors is a comparatively more proper approach to measure the market risks of portfolios (Glasserman et al., 2002; Johannes et al., 2009; Kamdem, 2009; Broda, 2012). Some other research developed the above market risk measurement models to measure credit portfolio risk based on the multivariate *t*-copula and estimate the tail probability of the credit portfolio loss distribution using the two-step IS technique (e.g., Kang and Shahabuddin, 2005; Kostadinov, 2006; Bassamboo et al., 2008; Chan and Kroese, 2010; Reitan and Aas, 2010). In this paper we adopt multivariate *t*-copula to capture the credit portfolio risk, where the risk factors have heavy-tailed distributions.

We show the conceptual framework of the proposed credit portfolio risk estimation in Figure 1. First, we calculate the default probability of each individual firm using the classical structural credit model, with the integration of six dimensions of data. Second, the multivariate *t*-copula being adopted to capture the credit portfolio risk is tested/estimated by normality test and nonparametric kernel density estimation; then, we integrate the multivariate *t*-copula into the classical structural credit model to calculate the joint default probability of the credit portfolio, which contains information on multiple dimensions of data. Third, we obtain the ES value of portfolio losses using the proposed three-step IS method for the *t*-copula credit portfolio risk measurement model. We further obtain the total provision for the risk of the portfolio using the achieved ES value.

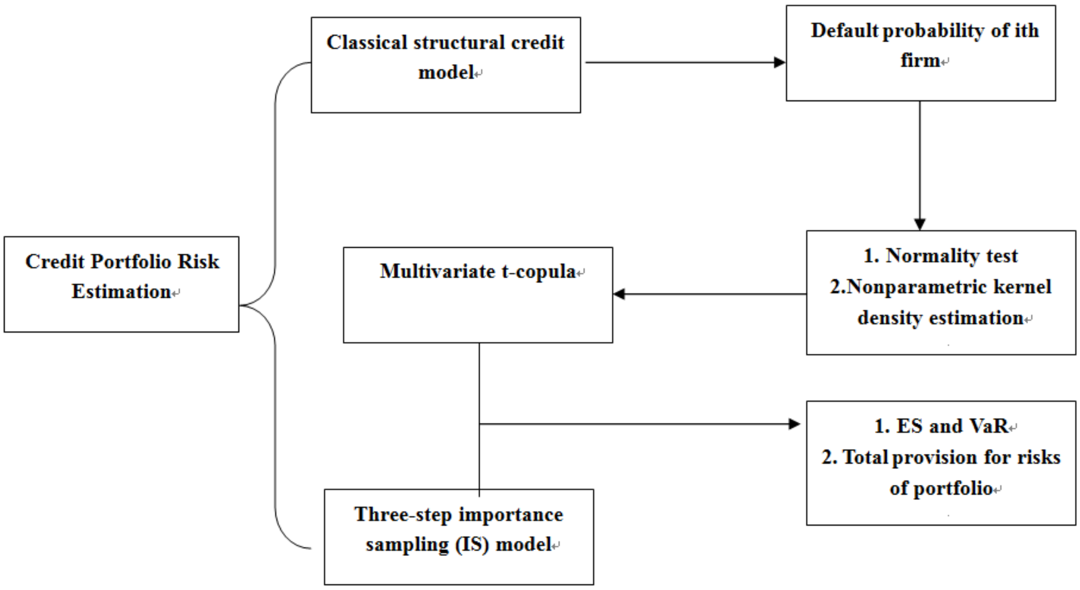

Figure 1 The conceptual framework of credit portfolio risk estimation

3 Calculation of default probability of each firm using the classical structural credit model

To measure the aggregated portfolio credit risk, we need to specify a model that can link defaults of several entities. Moreover, when measuring the credit portfolio risk, we face two fundamental problems: First, how to establish the correlation structure between the debtors' default probability, which we will address in the sequel. Second, how to integrate the credit portfolio and the economic environment, which reflects the real default expectation of the credit portfolio. In order to solve the second problem, we first solve the default probability of an individual firm based on the structural credit model. The structural credit model can be used to obtain the market-implied default probability of each firm with the explicit assumptions about the dynamics of the firm's assets and its volatility, and about the firm's debts and its leverage ratio, return on assets, and interest rate level. In other words, the structural credit model integrates the six dimensions of data, i.e., firm assets, volatility, debts, leverage ratio, return on assets, and interest rate level, to calculate the default probability of a firm. Obviously, the structural credit model possesses economic and intuitive appeals.

The basis of the classical structural credit model, dating back to Merton (1974) and Black and Scholes (1973), is that a firm's liabilities are contingent claims on the market value of the firm. The asset value of a firm is the ultimate source of indeterminacy driving credit risk. Therefore, pricing equity and credit risky debt reduce to pricing European options in the classical structural credit model.

Hillegeist et al. (2004) discovered that the structural credit model can provide significantly more information than either of two popular accounting-based measures of Altman's (1968) *Z*-score and Ohlson's (1980) *O*-score, which effectively summarizes publicly available information about the default probability of a firm. Vassalou and Xing (2004) found that the small-minus-big (SMB) and high-minus-low (HML) factors in Fama and French's three-factor model cannot replace the default risk factor, and that the risk factor obtained from the structural credit model is more effective. In practice, Moody's KMV model uses the structural credit model to predict the default probability of an individual firm.

We consider the classical Black-Scholes setting. Suppose a firm has a market value V, which denotes the expected discounted future cash flow of the firm. The firm is financed by equity and a zero coupon bond with a face value K and a maturity date T. A contractual obligation of the firm is to pay back the amount K to the bond investors at time T. If it fails to pay, the bondholders will have the right to take over the firm. Define V_T as a firm's market value on the maturity date T. To estimate the default probability, we suppose that changes in the value of the firm's assets over time follow the geometric Brownian motion, i.e., $\frac{dV_t}{V_t} = \mu dt + \sigma dW_t$, where $\mu \in R$ is the drift parameter, $\sigma > 0$ is the volatility parameter, and W_t is the standard Brownian motion. Suppose $m = \mu - \frac{1}{2}\sigma^2$. By Ito's lemma, we obtain $V_t = V_0 e^{mt+\sigma W_t}$, where $0 < t < T$.

Assuming that the firm defaults on the maturity date T, then the default probability $p(T)$ is given by

$$p(T) = P(V_T < K) = P(\sigma W_T < \log L - mT) = \Phi(\frac{\log L - mT}{\sigma T}), \quad (1)$$

where $L = \frac{K}{V_0}$ is the primary financial leverage ratio, and Φ is the standard normal cumulative distribution function (CDF).

Because pricing an equity and its credit risky debt are ascribed to pricing a European option, the equity value E_t of a firm is equal to a European call option on the firm's assets V with maturity T, strike price DB (i.e., DB_i is the debt of the i-th firm), and risk-free interest rate r. Supposing that the equity value E_t and its volatility σ^E are known, Jones et al. (1984) indicated that the market value of a firm's assets and its volatility can be derived by an option pricing formula. The market value of an asset and its volatility can usually be estimated through the asset's stock price, its stock price volatility, and the book value of its underlying debt.

In the classical structural model, a firm in the market, say, the *i*-th firm, defaults if its asset value is below the value of its debt, i.e.,

$$D_i = 1 \Leftrightarrow W_T^i < B_i , \quad (2)$$

where $W_T^i = \frac{\log(V_T^i / V_t^i) - m_i T}{\sigma_i}$, $B_i = \frac{\log(L_i) - m_i T}{\sigma_i}$, $m_i = \mu_i - \frac{1}{2}\sigma^2{}_i$, $L_i = \frac{K}{V_t^i} = \frac{DB_i}{V_t^i}$ is the financial leverage ratio, DB_i is the debt of the i-th firm, and D_i is the default indicator function for the i-th firm (1 if the i-th firm defaults; and 0, otherwise). If the time interval $T-t$ is equivalent to a unit time, W_T^i is the standardized asset returns of the i-th firm and B_i is the standardized book value of the debt of the i-th firm, which is also called the default distance.

Moreover, using Eq. (1), we can re-write Eq. (2) as follows:

$$p_i = P(W^i < B_i) = \Phi(B_i), \tag{3}$$

where Φ is the standard normal CDF. Once $W_T^i = \dfrac{\log(V_T^i / V_t^i) - m_i T}{\sigma_i}$ and $B_i = \dfrac{\log(L_i) - m_i T}{\sigma_i}$ are estimated, we can compute the default probability p_i using Eq. (3).

It is evident that we integrate the six dimensions of data of firm assets, volatility, debts, leverage ratio, return on assets, and interest rate level, to calculate the default probability of each firm using Eq. (3). In other words, we integrate six dimensions of data to obtain the default probability of each firm. Likewise, the dimension of the portfolio increasing can be dealt with in the same way using this data analytics technique. In the following empirical tests, we apply this data analytics technique to discover useful information about credit risk of each firm by exploiting pertinent data from different sources.

4 Default probability of the credit portfolio under the *t*-copula

Given that a large number of empirical studies have found that financial variables often show high kurtosis and heavy tails, the dependence between financial variables is generally extremal (Behr and Pötter, 2009; Jules, 2012; Schneider and Schweizer, 2015; Ankudinov et al., 2017; D'Amico and Petroni, 2018). In view of this, we use the *t*-copula model of credit risk to depict the extreme dependence of financial assets. In the *t*-copula model, the latent variables corresponding to the counterparties (risk factors) are supposed to have the multivariate *t*-distribution, instead of the multivariate normal distribution. Specifically, we use the multivariate *t*-distribution as the ratio of a multivariate normal and the square root of a scaled chi-squared random variable, which has been adopted in previous research (see, e.g., Kang and Shahabuddin, 2005; Kostadinov, 2006; Bassamboo et al., 2008; Chan and Kroese, 2010; Reitan and Aas, 2010).

We introduce the multi-factor linear model in the multivariate *t*-copula setting, which has been applied by Kang and Shahabuddin (2005), Kostadinov (2006), Bassamboo et al. (2008), Chan and Kroese (2010), and Reitan and Aas (2010). In this case, the standardized returns on asset W^i of the i-th firm can be parameterized by the multi-factor linear model

$$W^i = \sqrt{\frac{\nu}{V}}(a_i Z + b_i \varepsilon_i) = \sqrt{\frac{\nu}{V}}(\sum_{k=1}^{d} a_{ik} Z_k + b_i \varepsilon_i), \tag{4}$$

where Z is multivariate normal with the zero mean vector and the covariance matrix Σ, ε_i are independent and standard normal, $\mathrm{a}_i = (a_{i1}, \ldots a_{id})$ is a row vector of constant factor loadings for the i-th firm, b_i is the idiosyncratic factor loadings, and $V \sim \chi^2(\nu)$ is a chi-square distributed random variable with ν degrees of freedom. By Eq. (4), conditional upon both Z and V, we see that the conditional default probabilities are independent. Thus, we derive the conditional default probability of the i-th firm as follows:

$$p_i(Z,V) = P(W^i < B_i \mid Z,V) = \Phi\left(\frac{B_i - \sqrt{\frac{\nu}{V}} a_i Z}{\sqrt{\frac{\nu}{V}} b_i}\right). \tag{5}$$

Here, $p_i(Z,V)$ is the conditional default probability of the *i*-th firm and B_i is the standardized book value of the *i*-th firm, namely the standardized default threshold. Using the fact that the defaults conditional upon Z and V are independent, we have the joint default probability of the credit portfolio under the multivariate *t*-copula as

$$
\begin{aligned}
& p\left(D_1=1, D_2=1, \cdots, D_n=1\right) \\
& = E\left(p\left[D_1=1, D_2=1, \cdots, D_n=1 \mid Z, V\right]\right) \\
& = E\left(\prod_{i=1}^{n} p_i(Z, V)\right) \\
& = \int_{R^m} \prod_{i=1}^{n} \Phi\left(\frac{\sqrt{\frac{V}{\nu}} B_i - a_i Z}{b_i}\right) \phi_d(z; \Sigma) f_V dz
\end{aligned}
\tag{6}
$$

where $\phi_d(z;\sum)$ indicates the d-variate normal density function with the covariance matrix $\sum$ and f_V indicates the chi-square distributed density function with ν degrees of freedom.

It is evident that integrating multivariate *t*-copula into the classical structural credit model to calculate the joint default probability of the credit portfolio using Eq. (6). Namely, more than six dimensions of data are being integrated to obtain the joint default probability of the credit portfolio. In the following empirical test, we apply this data analytics technique, i.e., integrating multivariate *t*-copula into the classical structural credit model, to discover useful information about credit risk of portfolio including 60 listed Chinese firms by exploiting data from different sources.

5 Importance sampling for the credit portfolio risk measurement model under the *t*-copula

A critical factor in the credit portfolio risk measurement model is to depict the dependence among the counterparties. To facilitate presentation, we introduce the following notation:

n = number of counterparties to which the portfolio is exposed,

$D_i =$ the default indicator for the i-th firm (1 if the i-th firm defaults; and 0, otherwise),

$p_i =$ the marginal default probability that the i-th firm defaults,

$c_i =$ the loss owing to default to the i-th firm,

$L =$ the total loss from defaults.

Then the total loss L is given by $L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} c_i * D_i$. (7)

Focusing on the distribution of loss L from default over a fixed horizon, we wish to find the tail probability of the loss distribution L, especially in the case where the value of the threshold x is large and the event $\{L > x\}$ is rare, i.e., $P(L > x) = E[I(L > x)]$, (8)

where $I(.)$ is the indicator function (1 if $L > x$; and 0, otherwise).

To capture the dependence among the obligators (firms), we introduce the dependence among the default indicators D_1, D_2, …, D_n. In the *t*-copula model, dependence is introduced through the multivariate *t* vector (W^1, W^2, …, W^n) of the standardized returns on asset. Each default indicator is denoted as $D_i = I[W^i < B_i]$, where $i = 1, 2, \cdots, n$ and the standardized default threshold B_i is selected to match the marginal default probability p_i. Let t_ν be the cumulative distribution function (CDF) of the *t*-distribution with ν degrees of freedom. In this case, we need $B_i = t_\nu^{-1}(1 - p_i)$ to guarantee that $P(D_i = 1) = p_i$. A few studies also use a similar form, e.g., Kang and Shahabuddin (2005), Kostadinov (2006), Bassamboo et al. (2008), Chan and Kroese (2010), and Reitan and Aas (2010).

5.1 Monte Carlo simulation based on importance sampling

It is noted that applying Monte Carlo simulation based on importance sampling requires selecting a proper probability, which makes the rare event more likely to happen under different changes in the probability measure. Monte Carlo simulation is generally applied to estimate the tail probability of the loss function in Eq. (8). However, for a large value of x, very few samples actually have $L > x$, leading to the outcome that most samples are wasted because $I(L > x) = 0$.

The importance sampling (IS) technique is particularly well fit for simulating such a rare event. It matches the VaR problem with a small tail probability (see, e.g., Glasserman et al., 2000; Glasserman et al., 2002; Glasserman and Li, 2005; Grundkeb, 2009; Fernández et al., 2012; Xie et al., 2019). When using IS technique to evaluate a tail probability $P(L > x)$, a key consideration is the proper selection of a proposed default probability that makes the rare event $\{L > x\}$ more likely to happen. Specifically, in using the IS technique, we engender samples from the proposed default probability q_i instead of the original default probability p_i. Afterwards, the basic IS identity is

$$P(L > x) = \tilde{E}\left[I(L > x) \prod_{i=1}^{n} \left(\frac{p_i}{q_i} \right)^{D_i} \left(\frac{1-p_i}{1-q_i} \right)^{1-D_i} \right], \tag{9}$$

where $I(.)$ is the indicator function of the event in the braces, $\tilde{E}$ indicates that the expectation is based upon the new default probability q_i, and $\prod_{i=1}^{n} \left(\frac{p_i}{q_i} \right)^{D_i} \left(\frac{1-p_i}{1-q_i} \right)^{1-D_i}$ is the likelihood ratio (LR) relating the original distribution of (D_1 , D_2 , …, D_n) to the new one. Therefore, the term $I(L > x) \prod_{i=1}^{n} \left(\frac{p_i}{q_i} \right)^{D_i} \left(\frac{1-p_i}{1-q_i} \right)^{1-D_i}$ inside the expectation is an unbiased estimator of $P(L > x)$, with the default indicator sampled based on the new default probability.

5.2 Exponential twisting

In the following section we show how to obtain the likelihood ratio when applying the IS technique. To do so, we need to change the conditional marginal default probabilities from the original to the new default probabilities under exponential changes in the probability measure. Rather than raising the default probability discretionarily, we restrict the IS technique so that it corresponds to exponential twisting of the form L (expressed in Eq. (7)). Exponential twisting often appears in analysis of rare events and the corresponding IS technique (see, e.g., Huang and Shahabuddin, 2003). We apply IS through changing the conditional marginal default probabilities from p_i to exponentially twisted one $p_{i,\theta}$. Let θ be the exponential twisting parameter for all i and define

$$q_i = p_{i,\theta} = \frac{p_i \exp(\theta c_i)}{1+p_i\left(\exp(\theta c_i)\text{-}1\right)}. \quad (10)$$

Then, if $\theta > 0$, the new default probability $p_{i,\theta}$ after exponential twisting is greater than the original p_i. If $\theta{=}0$, the original default probability is invariant.

Given $D_i =$ the default indicator for the i-th firm (1 if the i-th firm defaults; and 0, otherwise) and the moment generating function of the 0-1 distribution just also $\left[1+p_i\left(\exp(\theta c_i)\text{-}1\right)\right]$, $\frac{p_i}{p_{i,\theta}} = \exp(-c_i\theta)\left[1+p_i\left(\exp(\theta c_i)\text{-}1\right)\right] = \exp(-c_i\theta)\exp\left(\varphi_i(\theta)\right)$, where $\varphi_i(\theta)$ is the logarithm of the moment generating function of a random variable that follows the binomial distribution c_iY_i, i.e., the cumulative moment generating function (CGF) of c_iY_i. $\frac{1-p_i}{1-p_{i,\theta}} = \exp\left(\varphi_i(\theta)\right)$ is the likelihood ratio in the absence of default. There is no $\exp(-c_i\theta)$ factor, because in that case there is no default, i.e., default loss $c_i = 0$.

Thus after exponential twisting by changing the default probability $p_{i,\theta}$, the LR of

$$\prod_{i=1}^{n}\left(\frac{p_i}{q_i}\right)^{D_i}\left(\frac{1-p_i}{1-q_i}\right)^{1-D_i} \text{ simplifies to } \prod_{i=1}^{m}\left(\frac{p_i}{p_{i,\theta}}\right)^{D_i}\left(\frac{1-p_i}{1-p_{i,\theta}}\right)^{1-D_i} = \exp(-\theta L + \psi(\theta)), \quad (11)$$

where $\psi(\theta) = \log E\left(\exp(\theta L)\right) = \sum_{i=1}^{n}\log(1+p_i(\exp(\theta c_i)-1))$ is the logarithm of the moment generating function of L, i.e., the cumulative moment generating function (CGF) of L. Furthermore, for any θ, the estimator $I(L>x)\exp(-\theta L+\psi(\theta))$ is an unbiased estimator of $P(L>x)$, with the total loss from defaults L sampled based on the new default probability $p_{i,\theta}$. Eq. (11) means that exponential twisting the default probability as in Eq. (10) is equal to using an exponential twisting of L itself.

5.3 Variance reduction analysis and choice of the exponential twisting parameter

In this section we explain why the variance of the estimator is reduced when the IS technique is applied. The exponential twisting parameter θ needs to be chosen while carrying out the probability measure change in the default probability. The variance of the estimator of $P(L>x)$ also decreases with such a probability measure change. In the following we address these issues.

To reduce the variance of the estimator of $P(L>x)$, we minimize its second moment under IS. The second moment of the unbiased estimator of $P(L>x)$ is

$$m_2(x,\theta)=E_\theta\left[\exp(-2\theta L+2\psi(\theta))I(L>x)\right]\le\exp(-2\theta x+2\psi(\theta)), \quad (12)$$

where E_θ indicates the expectation under IS with the exponential twisting parameter θ and the upper bound holds for all $\theta\ge 0$. While finding the value of the exponential twisting parameter θ is difficult, minimizing the upper bound in Eq. (12) is a simple matter. This amounts to minimizing $\psi(\theta)-\theta x$ over $\theta>0$. The logarithm of the moment generating function $\psi(\theta)$ is strictly convex and passes through the origin point, so the upper bound in Eq. (12) is minimized at

$$\theta_x=\begin{cases} solution\ \ to\ \ \psi'(\theta)=x, & x>\psi'(0) \\ 0 & x\le\psi'(0) \end{cases}. \quad (13)$$

The root of the above equation is easily solved numerically. The value θ_x determined by Eq. (12) has an additional interpretation that makes it appealing. The function $\psi(\theta)$ is the logarithm of the moment generating function (MGF) of the random variable L. By differentiating the function

$\psi(\theta)=\log E\left(\exp\left(\theta L\right)\right)$ and the definition of the exponential change of the probability measure, we obtain

$$\psi'(\theta) = E_\theta[L\exp(\theta L - \psi(\theta))\exp(-\theta L + \psi(\theta))] = E_\theta[L]. \tag{14}$$

According to Eq. (14), the expected value of $E_\theta[L]$ under the new probability measure is equal to x by Eq. (13). While x is in the tail of the original probability measure, it is near the centre of the new distribution when the IS technique is applied. In other words, the rare events $\{L>x\}$ are more probable to occur under the new probability measure P_θ and are no longer rare events. Furthermore, the variance of the estimator of $P\left(L>x\right)$ is reduced. Meanwhile, $I(L>x)\exp(-\theta L+\psi(\theta))$ is an unbiased estimator of $P\left(L>x\right)$, too.

5.4 Three-step importance sampling technique

We extend the two-step IS procedure to the three-step IS procedure for the *t*-copula credit portfolio risk measurement model with a view to achieving further variance reduction Three-step importance sampling can better capture the distributional property of credit portfolio loss, namely the tail probability, while the dependence structure of the multiple counter parties in the portfolio conforms to the *t*-distribution. The key to the three-step IS procedure, which will be used for the *t*-copula credit portfolio risk measurement, is to obtain the corresponding likelihood ratio in each step under exponential changes in the probability measure. This is essential to the importance sampling technique because we can calculate the tail probability of the loss distribution of the credit portfolio using the likelihood ratio in each step, and further obtain the total provision for risk in the algorithm in Section 6.2. Now we discuss how to achieve the corresponding likelihood ratio in each step when applying the three-step importance sampling technique.

We have used the first step of IS to effectuate probability measure change in the default probability. Now we apply the other two steps of IS, conditional on both the systematic risk factors Z and the chi-square distributed random variable V. Conditional on both Z and V, the default indicators D_1, D_2, …, D_n are mutually independent and the *i*-th firm has the conditional default probability $p_i(Z,V)$ as defined in Eq. (5). Applying $B_i = t_\nu^{-1}(p_i)$, we express this probability as

$$\begin{aligned} p_i(Z,V) &= P\left(D_i = 1 \middle| Z,V\right) \\ &= P\left(W^i < B_i \middle| Z,V\right) \\ &= \Phi\left(\frac{\sqrt{\frac{V}{\nu}}\, t_\nu^{-1}(p_i) - a_i Z}{b_i}\right). \end{aligned} \tag{15}$$

Therefore $\theta_x(Z,V)$ is solved first by substituting p_i for $p_i(Z,V)$ in Eq. (13), then the revised conditional default probability $p_{i,\theta_x}(Z,V)$ after exponential changes of the probability measure is determined by substituting θ_x for $\theta_x(Z,V)$, and by substituting p_i for $p_i(Z,V)$ in Eq. (10), i.e., Eq. (10) becomes $p_{i,\theta_x}(Z,V) = \dfrac{p_i(Z,V))e^{\theta_x(Z,V)c_i}}{1+p_i(Z,V)(e^{\theta_x(Z,V)c_i}-1)}$. (16)

By Eq. (11), we find the first likelihood ratio for the conditional IS default probability as

$$\ell_1 = \prod_{i=1}^{m} \left(\frac{p_i(Z,V)}{p_{i,\theta_x(Z,V)}(Z,V)}\right)^{D_i} \left(\frac{1-p_i(Z,V)}{1-p_{i,\theta_x(Z,V)}(Z,V)}\right)^{1-D_i} = \exp[-\theta_x(Z,V)L + \psi(\theta_x(Z,V))]. \tag{17}$$

The IS distribution of the systematic risk factors Z under the new probability measure is a multivariate normal distribution with a shifted mean vector μ (we will discuss how to find μ to enhance the effectiveness of the IS procedures in the next section) and a covariance matrix equal to the identity matrix I. In other words, the systematic risk factors Z after the probability measure change are sampled from $N(\mu, I)$. The corresponding second likelihood ratio for the systematic risk factors

Z under the new probability measure is

$$\ell_Z = \frac{(2\pi)^{-d/2}\exp(-\frac{1}{2}Z^T Z)}{(2\pi)^{-d/2}\exp\{-\frac{1}{2}(Z-\mu)^T(Z-\mu)\}} = \exp(-\mu^T Z + \mu^T\mu / 2). \quad (18)$$

Besides using IS to estimate the loss probability, we also need to show how to sample the random variable V after the probability measure change in the process of simulation. We now pay attention to the exponential changes of the probability measure for V. Let $f_\nu(V)$ be the chi-square distributed density function with ν degrees of freedom and α be the exponential twisting parameter. Thus, the density function $f_\nu(V)$ is changed to $f_\alpha(V)$ under the new probability measure, i.e.,

$$\frac{f_\nu(V)}{f_\alpha(V)} = \exp(-\alpha V + \psi_V(\alpha)) \quad (19)$$

where $\psi_V(\alpha)$ is the logarithm of the moment generating function of V.

Since the logarithm of the moment generating function is $\psi_V(\alpha) = \log[\phi_V(\alpha)]$ and the moment generating function is $\phi_V(\alpha) = (1-2\alpha)^{-\nu/2}$, we apply Eq. (19) to obtain Theorem 1 below.

Theorem 1 *If $f_\nu(V)$ is the chi-square distributed density function with ν degrees of freedom, α is the exponential twisting parameter, and the density function $f_\nu(V)$ is changed to $f_\alpha(V)$ after the exponential changes of the probability measure, then the density function $f_\alpha(V)$ is*

$$f_\alpha(V) = (\frac{2}{1-2\alpha})^{-\nu/2}\frac{V^{(\nu/2-1)}}{\Gamma(\nu/2)}\exp(-\frac{V}{2/(1-2\alpha)}). \quad (20)$$

The proof of Theorem 1 is given in Appendix 1, which demonstrates that we need to take the random variable from a transformed gamma distribution rather than the original chi-square distribution in the following algorithm. Eq. (20) indicates that $f_\alpha(V)$ is the gamma distributed density function with the scale parameter $\frac{2}{1-2\alpha}$ and shape parameter $\nu/2$, we need to sample the random variable V from this gamma distribution in the three-step IS technique.

Glasserman and Li (2005) developed the Zero Variance Measure method to determine the exponential twisting parameter α. We use their approach and obtain $\alpha = -c_n$, (21)

where $c_n = \frac{1}{2} n \|\gamma^*\|^2$, $\gamma^* = \min\{\|Z\|\}$, and Z are sampled from $N(\mu, I)$.

Furthermore, the corresponding second likelihood ratio for V after exponential changes of the probability measure by an amount $-c_n$ is given by

$$\ell_V = \frac{f(V)}{f_\alpha(V)} = \exp(-\alpha V + \psi_V(\alpha)) = \exp(c_n V)\, \phi_V(-c_n), \tag{22}$$

where $\phi_V(-c_n) = (1 + 2c_n)^{-\nu/2}$.

Therefore, in the IS technique, V (after exponential changes of the probability measure by an amount $-c_n$) is sampled from $\text{Gamma}(\nu/2, \frac{2}{1+2c_n})$, i.e., the gamma distribution with the scale parameter $\frac{2}{1+2c_n}$ and shape parameter $\nu/2$. Applying Eqs. (17), (18), and (22), we obtain

$$\begin{aligned} P(L>x) &= p(D_1=1, D_2=1, \cdots, D_n=1) \\ &= E\left(p\left[I(L>x) \middle| Z,V\right]\right) = E_{\theta_x(Z,V)}\left[I(L>x)\ell_1 \ell_Z \ell_V\right] \\ &= E_{\theta_x(Z,V)}\left\{I(L>x)FGH\right\}, \end{aligned} \tag{23}$$

where $F = \exp\left(-\theta_x(Z,V)L + \psi\left(\theta_x(Z,V)\right)\right)$, $G = \exp\left(-\mu^T Z + \mu^T \mu/2\right)$, and $H = \exp\left(c_n V + \psi_V(-c_n)\right)$.

Therefore, the term $I(L > x)\, e^{[-\theta_x(Z,V)L+\psi(\theta_x(Z,V))]} e^{-\mu^T Z+\mu^T \mu/2} e^{(c_n V+\psi_V(-c_n))}$ inside the expectation is an unbiased estimator of $P(L > x)$ with the exponential twisting parameter $\theta_x(Z,V)$ in the IS technique.

5.5 Mean-shifting method

To improve the effectiveness of the IS technique, we now discuss how to obtain the multivariate normal distribution shifted mean vector μ under exponential changes in the probability measure. By the Variance Decomposition Theorem, for any estimator $\hat{p}_x$ of $P(L > x)$, we have

$$Var(\hat{p}_x) = E(Var[\hat{p}_x \mid Z,V]) + Var(E[\hat{p}_x \mid Z,V]). \tag{24}$$

Conditional on Z and V, the firms are independent, and we know that minimizing the upper bound on the second moment of the IS estimator of the tail probability in Eq. (12) makes $Var[\hat{p}_x \mid Z,V]$ minimum or small when applying IS to Z and V. Therefore, in order to make the variance $Var(\hat{p}_x)$ minimum or small, we should pay attention to the second term $Var(E[\hat{p}_x \mid Z,V])$ in Eq. (24). Given that $\hat{p}_x$ is the estimator of $P(L > x)$ and $E[\hat{p}_x \mid Z,V] = P(L > x \mid Z,V)$, we have

$$P(L > x) = \iint I(L > x \mid Z,V) f_Z(Z) f_V(V) dZdV, \tag{25}$$

where $f_Z(Z)$ is the probability density function of the systematic risk factors Z and $f_V(V)$ is the probability density function of the chi-square distributed random variable V. This implies that the IS distributions for Z and V should reduce the variance in the integral of $P(L > x \mid Z,V)$ against the densities of Z and V.

The zero variance IS distribution for the multi-factor problem would sample the systematic risk factors Z from the probability density proportional to what is inside the integral in Eq. (25). According to Glasserman and Li (2005), Kang and Shahabuddin (2005), and Reitan and Aas (2010), sampling this probability density is generally impracticable. On the contrary, they suggested using a normal density with the same method as the optimal probability density. This method occurs at the solution to the optimization problem $\max_Z P(L > x \mid Z,V) \exp\left(-Z^T Z/2\right) \frac{1}{2^{\nu/2}\Gamma(\nu/2)} V^{\nu/2-1} e^{-V/2}$, (26) which is also the mean of the approximating normal distribution.

It is difficult to obtain the exact solution of Eq. (26). There are several methods to approximate the solution (see, e.g., Kang and Shahabuddin, 2005; Kostadinov, 2006; Bassamboo et al., 2008; Chan and Kroese, 2010; Reitan and Aas, 2010). However, these methods have slow convergence rates. To solve this problem, we use the Levenberg-Marquardt algorithm, a nonlinear optimization technique, to obtain the optimal mean-shift vector. The algorithm proposed by Levenberg (1944) and Marquardt (1963) has the combined merits of the Gauss-Newton algorithm and the gradient method.

5.6 Expected Shortfall

In this section we discuss how ES and VaR are closely related to each other. It is well known that VaR is not a coherent risk measure since it lacks the property of subadditivity when the risk factors have heavy-tailed distributions. In addition, VaR is sensitive to changes in the significance level. This leads to a lack of the numerical stability needed for bank management decisions (Frey and McNeil, 2002; Kalkbrener et al., 2004; Reitan and Aas, 2010).

In view of the drawbacks of VaR, we apply expected shortfall (ES), i.e., the expected excess loss, given that there are portfolio losses, to measure credit portfolio risk. Acerbi and Tasche (2002) defined ES as follows:

Definition 5.1. (Expected shortfall). Given an integrable random variable L, the ES at confidence level α is given by

$$ES_\alpha(L) = \frac{1}{1-\alpha}\left(E[LI_{L\geq VaR_\alpha(L)}] + VaR_\alpha(L)(1-\alpha-P[L\geq VaR_\alpha(L)])\right), \tag{27}$$

where $VaR_\alpha(L) = \inf\{l \in \mathbb{R}, P(L > l) \leq 1-\alpha\}$. For a random variable L with a continuous distribution, the second term in Eq. (27) disappears and Eq. (27) reduces to

$$ES_\alpha(L) = \frac{E[LI_{L \geq VaR_\alpha(L)}]}{1-\alpha} = E[L|L \geq VaR_\alpha(L)] = \frac{E[L; L \geq VaR_\alpha(L)]}{P(L \geq VaR_\alpha(L))}. \quad (28)$$

ES is a coherent risk measure and possesses the subadditivity property, which reflects the idea that risk can be reduced by diversification, a time-honoured principle in Finance and Economics (see Acerbi and Tasche, 2002, for a concrete proof). It has become an increasingly popular risk measure in the finance industry.

While VaR is sensitive to changes in the significance level, ES does not fluctuate drastically with changes in the significance level. Since VaR and ES are generally based on the same estimated loss distribution, ES is closely related to VaR according to Eq. (27) and Eq. (28) when ES is applied to compute credit portfolio risk. In fact, the two measures complement each other (Basel Committee on Banking Supervision, 2009). Thus, we apply VaR and ES together to measure credit portfolio risk based on the same estimated loss distribution.

6 Calculation of the total provision for risk

6.1 Total provision for risk

Decision makers need to make operational decisions based on their assessment of the risk of the firm in a specific industry. In Sections 3 and 4, we demonstrated how to estimate the credit portfolio risk with the proposed model. In this section, we illustrate how to calculate the total provision for risk.

Based on our proposed model, the decision maker can effectively measure the comprehensive risk of the portfolio position and obtain ES. Then the decision maker can effectively determine the operational decisions on the total provision for risk with consideration of credit risk and market risk, preparing the risk loss reserve in a timely and adequate manner. Specifically, the Basel III Accords

(2011) require the preparation of the total provision for risk, which consists of the provision for credit risk and the provision for market risk. Based on the Basel III Accords, the provision for credit risk should be at least equal to 8% of the bank's risk-weighted assets. The provision for market risk commonly equals the former ES, or the last 60 trading days' average ES multiplied by a multiplier factor k that is not less than 3. It follows that

$$TRC = CRC + MRC = 8\% \sum_i w_i A_i + \max\left(k \frac{1}{60} \sum_{i=1}^{60} ES_{t-i}, ES_{t-1}\right), \tag{29}$$

where TRC is the total provision for risk, CRC is the provision for credit risk, MRC is the provision for market risk, A_i is the value of the i-th asset, w_i is the weight of A_i, and the penalty factor k equals 3 (the lower limit value of the penalty factor k is 3 according to the Basel III Accords, for convenience of calculation, we set the penalty factor k as 3 in this paper). Eq. (29) shows that the estimated ES value directly affects the operational decisions of the total provision for risk. This relationship highlights the importance of estimating the credit portfolio risk with an efficient and effective method.

It is evident that on the basis of achieving an accurate estimation ES using the proposed three-step IS method for the t-copula credit portfolio risk measurement model, the paper further integrates the achieved ES value to calculate the total provision for risk using Eq. (29). Thus, the obtained total provision for risk reacting in a timely manner to economic and environmental changes is to be obtained by synergizing the classical structural credit model, the integrating multivariate t-copula into the classical structural credit model, and the proposed three-step importance sampling (IS) model. In the following empirical test, the eleven dimensions data are being integrated together to calculate the joint default probability of credit portfolio and obtain the total provision for risk for adapting to economic and environmental changes. This also further illustrates that this data analytics technique is useful to extract and utilize useful information about credit portfolio risk by exploiting data from different sources.

6.2 Algorithm to calculate the credit portfolio risk under *t*-copula condition

In order to calculate the credit portfolio risk under *t*-copula condition, we provide a formal algorithm to calculate the credit portfolio risk under the *t*-copula condition as follows:

Part 1 Parameter estimation based on classical structural credit model and multivariate *t*-copula

1) Calculate the individual default probability of the 60 listed firms based on the **classical structural credit model**.

2) Utilizing normality test and nonparametric kernel density estimation to test if each individual default probability of the 60 listed firms conforms to the *t*-distribution, and choose the appropriate copula function by tail coefficient of correlation.

3) The systematic risk factor loadings row vector for the *i*-th firm a_i, the idiosyncratic risk factor loadings for the *i*-th obligator (firm) b_i, and the ν degrees of freedom have been estimated based on the data of the 60 listed firms and the corresponding data on treasury yields in China, Shanghai real estate index, retail sales index, and financial sector index spanning the same period.

4) In applying the Levenberg-Marquardt algorithm, we need to estimate the mean-shift vector μ for Z after the probability measure change. Solve the shifted mean vector μ and the degree of freedom ν of Gamma under the condition of

$$\max_z P(L>x\,|\,Z=z,V=v)\exp\left(-Z^T Z/2\right)\frac{1}{2^{\nu/2}\Gamma(\nu/2)}v^{\nu/2-1}e^{-v/2}.$$

Part 2 Formal algorithm of the IS procedure for calculating credit portfolio risk based on the *t*-copula

Step 1. Sample Z from $N(\mu, I)$.

Step 2. Sample V from $\text{Gamma}(\nu / 2, \frac{2}{1+2c_n})$, i.e., the gamma distribution with the shape parameter $\nu/2$ and scale parameter $\frac{2}{1+2c_n}$, where $c_n = \frac{1}{2} n \|\gamma^*\|^2$ and $\gamma^* = \min\{\|Z\|\}$.

Step 3. Calculate the conditional default probability $p_i(Z,V) = \Phi\left(\frac{\sqrt{\frac{V}{\nu}} t_\nu^{-1}(p_i) - a_i Z}{b_i}\right)$, given p_i, a_i, b_i, and ν.

Step 4. Obtain the exponential twisting parameter $\theta_x(Z,V)$ through numerically solving Eq. (13).

Step 5. Calculate the twisted conditional default probability

$$p_{i,\theta_x}(Z,V) = \frac{p_i(Z,V))e^{\theta_x(Z,V)c_i}}{1 + p_i(Z,V)(e^{\theta_x(Z,V)c_i} - 1)}, \quad i = 1,2,\cdots,n.$$

Step 6. Calculate $W^i = \sqrt{\frac{\nu}{V}}(a_i Z + b_i \varepsilon_i)$ and $B_i = t_\nu^{-1}(p_{i,\theta_x}(Z,V))$, $i = 1,2,\cdots,n$. Note: Sample ε_i is from $N(0,1)$.

Step 7. Generate the default indicators $D_i = I[W^i > B_i]$ from the twisted conditional default probability $p_{i,\theta_x}(Z,V)$, $i = 1,2,\cdots,n$.

Step 8. Calculate $\psi(\theta_x(Z,V)) = \sum_{i=1}^{n} \log\left[1 + p_{i,\theta_x}(Z,V)\left[\exp\left(\theta_x(Z,V)c_i\right) - 1\right]\right]$ and $\psi_V(\text{-c}_n) = \frac{-\nu}{2} \log\left[1 + 2c_n\right]$.

Step 9. Compute the loss $L = \sum_{i=1}^{n} L_i = \sum_{i=1}^{n} c_i * D_i$.

Step 10. Calculate the likelihood ratios $\ell_1 = \exp[-\theta_x(Z,V)L + \psi(\theta_x(Z,V))]$, ℓ_Z

$= \exp(-\mu^T Z + \mu^T \mu / 2)$, and $\ell_V = \exp(c_n V + \psi_V(-c_n))$ by Eq. (17), Eq. (18), and Eq.(22), respectively.

Step 11. Return the estimator $I(L > x)FGH$. Note: $F = \exp\left(-\theta_x(Z,V)L + \psi\left(\theta_x(Z,V)\right)\right)$, $G = \exp\left(-\mu^T Z + \mu^T \mu / 2\right)$, and $H = \exp\left(c_n V + \psi_V(-c_n)\right)$.

Step 12. Obtain the estimator $\hat{p}$ of the tail probability $\frac{1}{N}\sum_{j=1}^{N}\left[I\left(L^{(j)} > x\right)F^{(j)}G^{(j)}H^{(j)}\right]$ through repeating Steps 1 to 11 N times (N is the number of simulation runs). Note: $F^{(j)} = \exp\left(-\theta_x\left(Z^{(j)},V^{(j)}\right)L + \psi\left(\theta_x\left(Z^{(j)},V^{(j)}\right)\right)\right)$, $G^{(j)} = \exp\left(-\mu^T Z^{(j)} + \mu^T \mu / 2\right)$, and $H^{(j)} = \exp\left(c_n^{(j)} V^{(j)} + \psi_V\left(-c_n^{(j)}\right)\right)$.

Step 13. Calculate the estimator $\hat{m}_2(x, \theta_x)$ of the second moment under IS $\frac{1}{N}\sum_{j=1}^{N}\left[I\left(L^{(j)} > x\right)\left(F^{(j)}\right)^2\left(G^{(j)}\right)^2\left(H^{(j)}\right)^2\right]$ through Eq. (12). Note: $\left(F^{(j)}\right)^2 = \exp\left(-2\theta_x\left(Z^{(j)},V^{(j)}\right)L + 2\psi\left(\theta_x\left(Z^{(j)},V^{(j)}\right)\right)\right)$, $\left(G^{(j)}\right)^2 = \exp\left(-2\mu^T Z^{(j)} + \mu^T \mu\right)$, and $\left(H^{(j)}\right)^2 = \exp\left(2c_n^{(j)} V^{(j)} + 2\psi_V\left(-c_n^{(j)}\right)\right)$.

Step 14. Estimate $E[LI_{L \geq VaR_\alpha(L)}]$.

Step 15. Return the estimator $\hat{ES}_\alpha(L) = \frac{E[LI_{L \geq VaR_\alpha(L)}]}{\hat{p}}$.

Step 16. Repeat Step1 to Step 15 when i ranges from 1 to 60 and estimate

$$TRC = CRC + MRC = 8\%\sum_i w_i A_i + \max\left(k\frac{1}{60}\sum_{i=1}^{60} ES_{t-i}, ES_{t-1}\right).$$

The above concrete steps of the algorithm to calculate the credit portfolio risk under the *t*-copula condition can be used to clarify the computational process.

7 Empirical tests

Most recent studies on estimating portfolio credit risk resort to numerical computation. However, given the inherent relationship between dependent default and asset returns in the classical structural model, we empirically examine the data of 60 listed Chinese firms. We apply our proposed model to three industries, namely the real estate industry, retail industry, and finance and insurance industry. They are classified as industries in the Shanghai Stock Exchange in China. We choose these three industries for three reasons. First, the three industries account for a large portion of the Gross Domestic Product (GDP), so they can effectively reflect the state of the national economy. Second, the three industries have their own industry indices, which contain systemic risk information. The information can be used as the values of the systematic risk factors. Third, the industry indices and their information are accessible to us. Particularly, we select 60 listed Chinese firms, including 25 firms from the real estate industry (serial numbers 1 to 25), 20 firms from the retail industry (serial numbers 26 to 45), and 15 firms from the finance and insurance industry (serial numbers 46 to 60) listed in the Shanghai Stock Exchange, as shown in Appendix 2. We source the data of the 60 firms from the China Stock Market & Accounting Research (CSMAR) database covering the period from 4 January 2012 to 7 June 2013, excluding the H shares, overseas listed stocks, and suspended stocks in the period. We also gather the corresponding data on treasury yields in China, Shanghai real estate index, retail sales index, and financial sector index (the value of the systematic risk factors) spanning the same period from the Flush iFind financial database.

During this period, the real estate industry in China faced a new round of regulation where five new national real estate regulation policies were promulgated on 20 February 2013 in China, which centred on home buying restrictions and loan limitations. The five new national real estate regulation policies are to perfect the responsibility system for stabilizing house prices, resolutely curbing speculative investment in housing purchases, increasing the supply of general commodity housing and land, speeding up the planning and construction of secure residence projects, and strengthening market supervision. Since the beginning of real estate market regulation and control in December 2009, the policy has gone through four upgrades, namely, the “National 11” in January 2010, the “National Ten” in April, the “9.29 New deal” in September, the eight new national real estate regulation policies in January 2011, and the five new national real estate regulation policies issued on 20 February, 2013. The focus of the series policies of real estate market regulation and control is on home buying restrictions and loan limitations. This means that the credit portfolio including 25 listed Chinese firms in the real estate industry will be affected by the stringent policies of real estate market regulation and control, and the default risk of the 25 listed Chinese firms in the real estate industry will increase.

We apply the proposed *t*-copula credit portfolio risk measurement model to compute the tail probabilities of the credit portfolio loss distributions of the 60 listed Chinese firms and their corresponding VaR and ES, obtaining large variance reductions. We also check the changes in credit portfolio risk in different industries corresponding to the new national real estate regulation policies promulgated in China in the same period to test the model’s ability to produce results that agree with the historical events that transpired.

7.1 Calculation of default probability of each firm

Since the market value of a firm’s assets and its volatility can be derived by an option pricing formula, the market value of an asset and its volatility can usually be estimated through the asset’s stock price, its stock price volatility, and the book value of its underlying debt. Using the classical structural model in Section 3, we calculate the market value of each listed firm in each day in the corresponding period. Based on the market values of the firms, we further use Eq. (2) and Eq. (3) to find the standardized asset returns of each firm $W_T^i = \frac{\log(V_T^i / V_t^i) - m_i T}{\sigma_i}$ and the corresponding default distance $B_i = \frac{\log(L_i) - m_i T}{\sigma_i}$. Using Eq. (3), we can obtain 60 default probabilities of the 60 firms in each day, 382 default probabilities of each firm in the corresponding period, 9,550 default probabilities of the 25 firms in the real estate industry, 7,640 default probabilities of the 20 firms in the retail industry, and 5,775 default probabilities of the 15 firms in the finance and insurance industry in the corresponding period. Thus, we can further obtain the three average default probabilities in the real estate industry, in the retail industry, and in the finance and insurance industry. We plot the daily average default probability in each industry in Figures 2-4, and give the overall average default probability in each industry in Table 1.

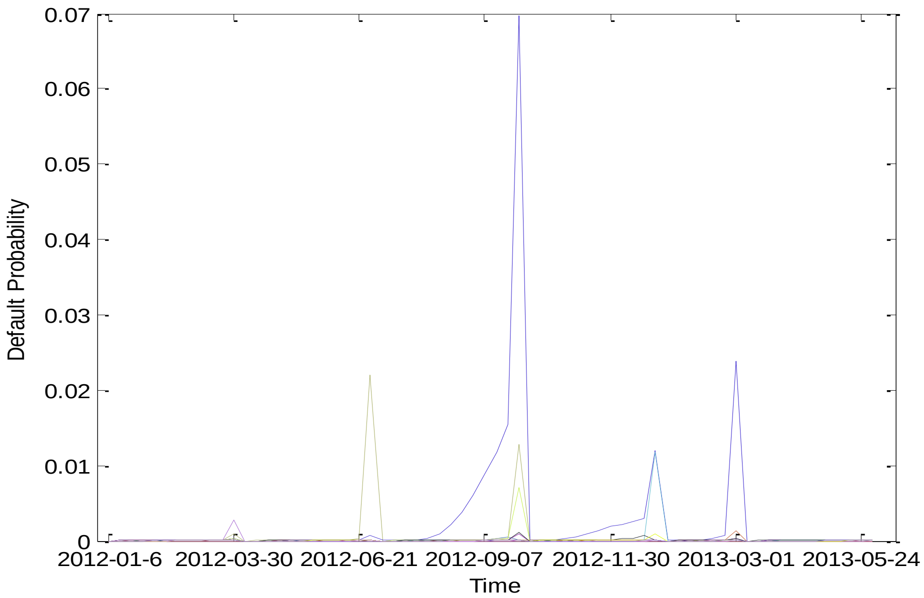

Figure 2 Trend of the default probability of each firm in the real estate industry

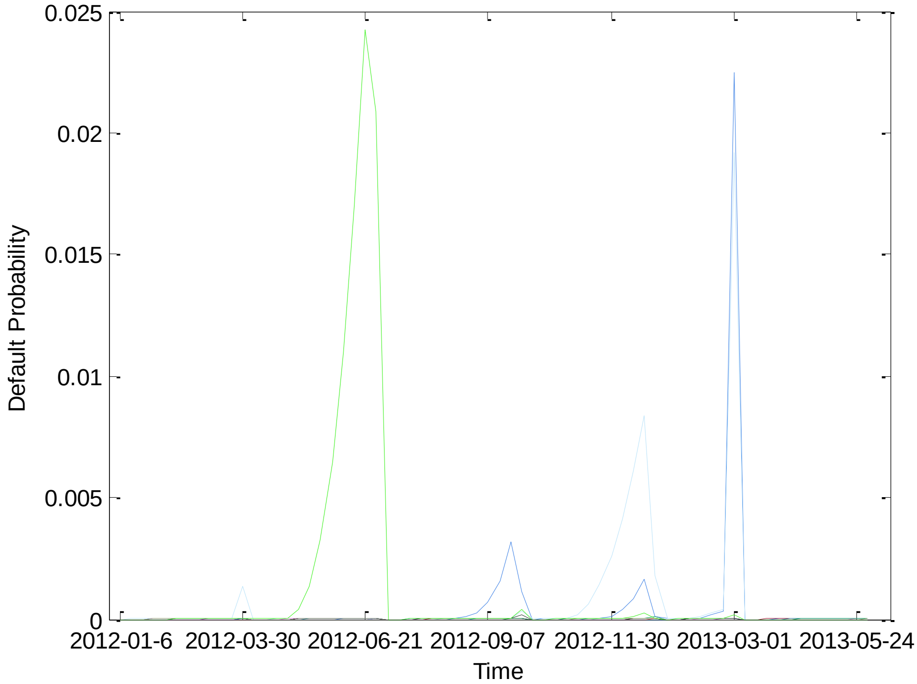

Figure 3 Trend of the default probability of each firm in the retail industry

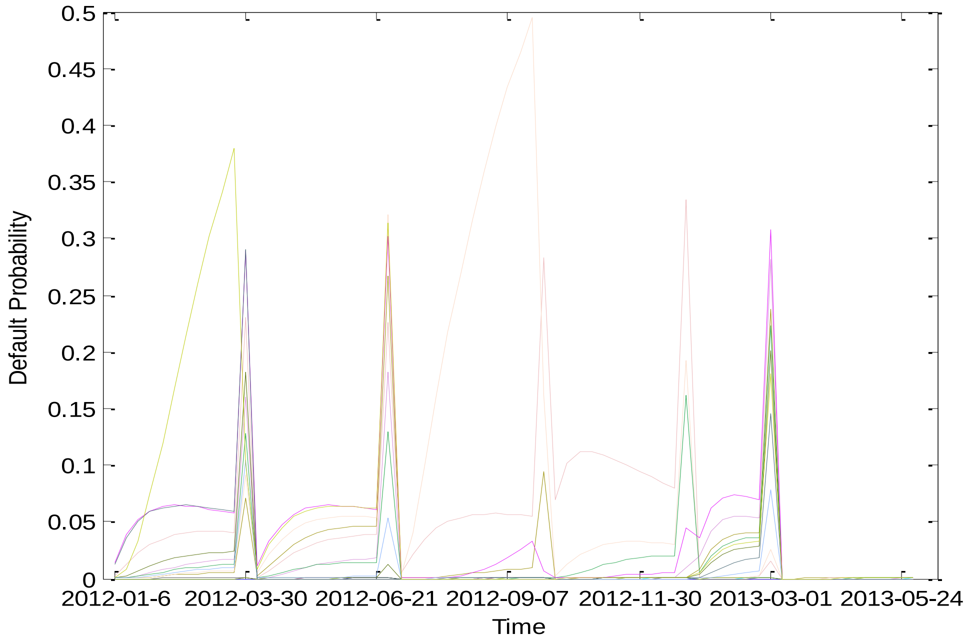

Figure 4 Trend of the default probability of each firm in the finance and insurance industry

Table 1 The average default probabilities of the firms in the three industries

	Finance and insurance firms	Real estate firms	Retail firms
Average default probabilities	0.031	0.0009	0.00002

Figures 2-4 and Table 1 show that the average default probability in finance and insurance firms is larger than those in the retail firms and real estate firms in general. This is related to the fact that banks rely on deposits to maintain their operations. Currently, the asset-liability ratios of banks are generally up to 90 percent and that of the insurance industry is around 80 percent. So the ratios are much higher than those of the real estate and retail industries.

Although the ratio of asset-liabilities in the real estate industry is high and its cash flow is generally tighter, we only consider assets and liabilities, i.e., we do not examine the cash flows of the firms and other related factors in this paper. Also, we consider only the data disclosed in the financial statements of the firms and do not consider off-balance sheet businesses such as implicit liabilities. Therefore, the overall risk of the finance and insurance industry is relatively large when the house prices remain high and no inventory valuation loss is incurred in the real estate business.

It is evident that, in the process of the computational experiments, using the classical structural credit model, we integrate the six dimensions of data embracing firm assets, volatility, debts, leverage ratio, return on assets, and interest rate level, to calculate the default probability of each listed firm on each day in the corresponding period. Specifically, we input a total of 137,520 data using the classical structural credit model to generate 60 values of the default probability of each firm on each day. Likewise, the dimension of the portfolio increasing can be dealt with in the same way using this data analytics technique. This illustrates that data analytics technique is useful to discover information about credit risk of individual firm by exploiting data from different sources.

7.2 The credit portfolio risk measurement model based on the *t*-copula

7.2.1 Normality test

It is essential to perform a goodness-of-fit test on the copula function to ascertain its viability. Whether the copula function can fully describe the structure of the dependence between financial variables depends on whether the selected copula model is proper and reasonable. We determine the appropriate copula model by conducting a series of normality tests on the pertinent financial variables, such as the KS-test, JB-test, and Lillie-test. Using the classical structural credit model that pricing equity and credit risky debt reduces to pricing European options, we calculate the market value of each listed firm on each day in the corresponding period. Furthermore, by standardizing the asset returns W_t^i of the i-th firm using Eq. (1), we perform a series of correlation analysis on the standardized asset returns to test whether they conform to the normal distribution. We report the interval values of the results in Table 2 and provide the details in Appendix 3.

Table 2 Normality test of standardized asset returns

	Statistics		**KS-test**		**JB-test**		**Lillie-test**	
Interval value of standardized asset returns	Skewness	Kurtosis	h	P	h	P	h	p
Maximum value	3.28	20.1	1	0.001818	1	0.001	1	0.001
Minimum value	-1.66	3.1	1	0.0000476	1	0.001	1	0.001

The results of the normality test show that, according to the KS-test, JB-test, and Lillie-test, all the values of the test statistic H for the idiosyncratic risk of each firm is 1. This means that the standardized asset returns do not follow the normal distribution, so we exclude the possibility that the tail is uncorrelated.

In view of the test statistics, and the corresponding skewness and kurtosis values, we may conclude that the standardized asset returns follow a heavy-tailed distribution, so the normal copula is not suitable for capturing the structure of the dependence between the assets for calculating credit portfolio risk. Since the normality test cannot reveal the distribution of the standardized asset returns, we use the nonparametric kernel density estimation to fit its distribution.

7.2.2 Nonparametric kernel density estimation

Observing the shape of the sample distribution is beneficial to further determining the form of the copula. Based on the estimated results, we use the distribution curves of the standardized asset returns due to Poly Real Estate in the real estate industry, Minmetals Development in the retail industry, and Southwest Securities in the finance and insurance industry to address our problem.

Figures 5-7 show that the nonparametric kernel density estimation can better fit the distribution of the sample, so we use this estimation to fit the distribution of the standardized asset returns of each firm.

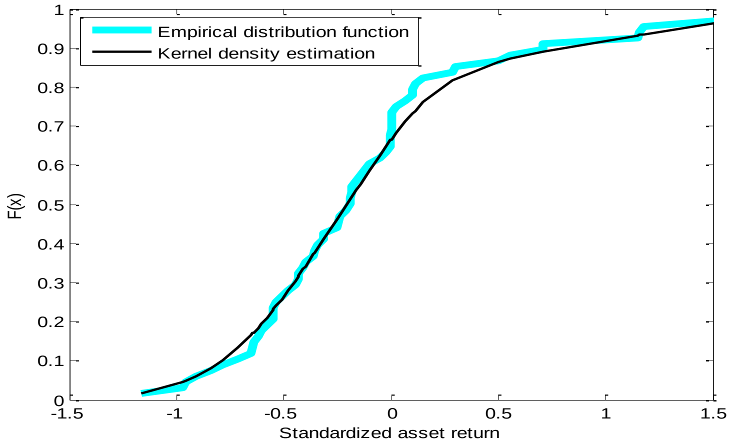

Figure 5 Distribution curves due to Poly Real Estate in the real estate industry

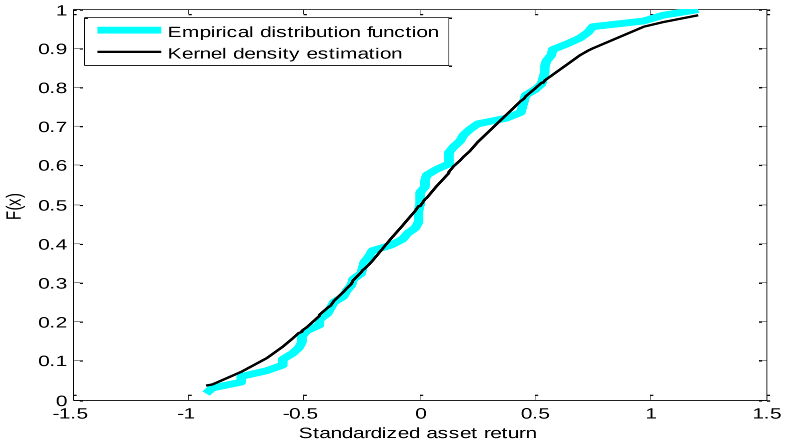

Figure 6 Distribution curves due to Minmetals Development in the retail industry

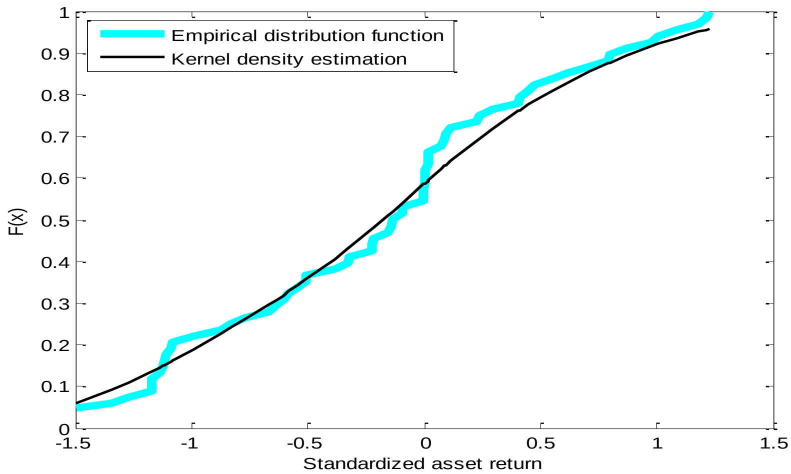

Figure7 Distribution curves due to Southwest Securities in the finance and insurance industry

7.2.3 Selecting an appropriate copula for calculating credit portfolio risk

On the premise that the marginal distribution of the standardized asset returns of each firm is known, we choose the appropriate copula function to construct the dependence between assets. Considering whether or not its distribution is symmetrical, we calculate the tail correlation coefficient between assets by identifying whether or not the tail correlation coefficient is 0 to determine the form of the copula function.

Figure 8 shows that the tail coefficient of correlation between assets is not zero, which shows obvious tail dependence between assets. It also shows that the upper tail and lower tail correlation coefficients are equal, which establishes that its distribution is symmetrical. Thus, it is reasonable to assume that the distribution of the standardized asset returns follows the *t*-distribution, and choosing the *t*-copula can better capture the tail dependence between assets.

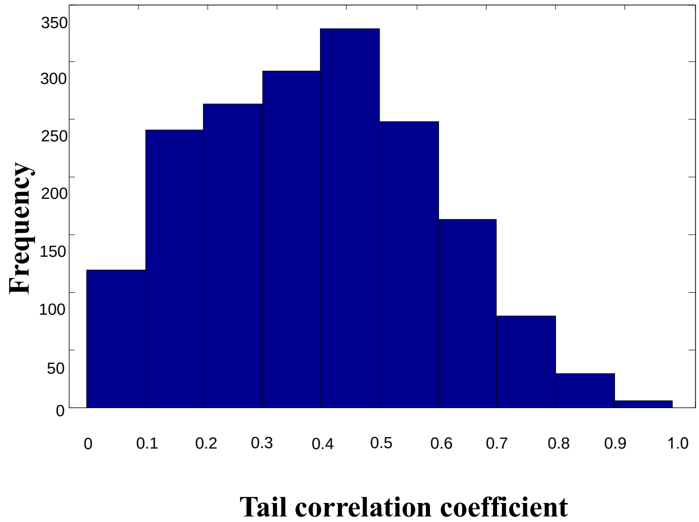

Figure 8 Histogram of the tail correlation coefficient

7.2.4 Parameter estimation for risk factor loadings

Since the distribution of the standardized asset returns shows obvious tail dependence between assets, we select the *t*-copula to capture the tail dependence. We use Eq. (4) that combines the various systematic and idiosyncratic risk factors to calculate credit portfolio risk. First, we estimate the various risk factor loadings as follows:

$$W^i = \sqrt{\frac{\nu}{V}}(a_i Z + b_i \varepsilon_i) = \sqrt{\frac{\nu}{V}}\left(\sum_{k=1}^{n} a_{ik} Z_k + b_i \varepsilon_i\right), \text{ and } \overline{\alpha_{i,k}} = \sqrt{\frac{\nu}{V}} a_{ik}, \overline{b_i} = \sqrt{\frac{\nu}{V}} b_i, k = 1, 2, 3, 4.$$

We select the corresponding data on treasury yields in China, Shanghai real estate index, retail sales index, and financial sector index as the values of the systematic risk factors, and then standardize the data. For each standardized asset returns of a listed firm we need to find the corresponding factor loadings $\overline{a_{i1}}, \overline{a_{i2}}, \overline{a_{i3}}, \overline{a_{i4}}$ of the systematic risk factors and the factor loading $\overline{b_i}$ of the idiosyncratic risk factor. Through the copula fit function, we obtain that the optimal degree of freedom is 4. We also estimate the factor loadings of the systematic and idiosyncratic risk factors by nonlinear estimation in Appendix 4.

7.2.5 Estimating the joint default probability of the credit portfolio under the *t*-copula

The analysis in Section 6.1.4 establishes that the *t*-copula model with a degree of freedom of 4 can capture the tail dependence between assets through the copula fit function. Thus, using the factor loading estimators $\overline{a_{i1}}, \overline{a_{i2}}, \overline{a_{i3}}, \overline{a_{i4}}$ and $\overline{b_i}$, we use Eq. (6) to estimate the historical joint default probability of the credit portfolio under the *t*-copula and show the results in Fig. 9 (Note: The order of magnitude of the vertical axis of the default probability is 10^{-7}).

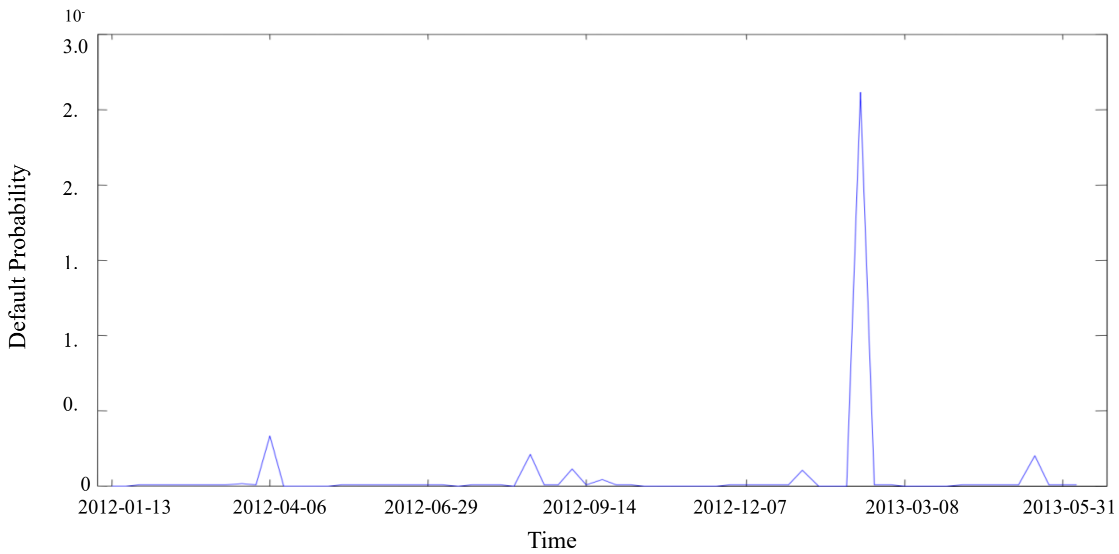

Figure 9 Joint default probability of the credit portfolio

As Figure 9 shows, from January 2012 to June 2013, the joint default probability of the credit portfolio was small as a whole; while from February 2013 to March 2013, the joint default probability fluctuated significantly. There are two reasons for this. First, industries in China were facing a downward domestic economic environment. This means that the whole credit portfolio including 60 listed Chinese firms was affected by the downward domestic economic environment, and the default probability of the whole credit portfolio would increase during this period. Second, the real estate industry in China was facing a new round of regulation, i.e., China promulgated five new national real estate regulatory policies on 20 February 2013, stipulating home buying restrictions and loan limitations. This means that the credit portfolio including 25 listed Chinese firms in the real estate industry would be affected by the stringent policies of real estate market regulation and control, and the default probability of the 25 listed Chinese firms in the real estate industry will increase with five new national real estate regulatory policies being promulgated. Meanwhile, the domestic consumption in China was weak during this period. This means that the credit portfolio including 20 listed Chinese firms in the retail industry would be affected by the weak domestic consumption, and the default probability of the 20 listed Chinese firms in the retail industry would increase. Later on, the impact of the real estate regulation was digested by the market and housing prices stopped falling and became stable, and domestic consumption rebounded. Therefore, the joint default probability of the credit portfolio remained low. This means that integrating multivariate *t*-copula into the classical structural credit model can detect the credit risk of the portfolio including 60 listed Chinese firms covering the period from 4 January 2012 to 7 June 2013.

Obviously, in the process of the computational experiments, using the integrating multivariate *t*-copula in the classical structural credit model, the eleven dimensions of data are being integrated together, i.e., the five dimensions of data embracing treasury yields in China, Shanghai real estate index, retail sales index, financial sector index, and return on assets, as the values of the systematic risk factors is being integrated with the original six dimensions of data embracing firm assets, volatility, debts, leverage ratio, return on assets, and interest rate level, collected from 60 listed Chinese firms covering the period from 4 January 2012 to 7 June 2013. We integrate the eleven dimensions of data to compute the joint default probability of credit portfolio including 60 listed Chinese firms on each day in the corresponding period. In other words, we input a total of 252,120 data to generate 382 joint default probabilities of the credit portfolios each day including 60 listed Chinese firms covering the period from 4 January 2012 to 7 June 2013. Likewise, the dimension of the portfolio increasing can be managed in the same way using this data analytics technique, i.e., integrating multivariate *t*-copula with the classical structural credit model. This further illustrates that data analytics techniques are useful to discover information about credit portfolio risk by exploiting pertinent data from different sources.

7.3 Estimator of credit portfolio risk

We now illustrate the performance of our three-step IS procedure based on the multivariable *t*-distribution. Making an accurate calculation of the tail probability of the credit portfolio loss distribution or the corresponding loss is crucial for risk management. A good estimation of the risk value facilitates financial institutions to keep adequate loss reserves to protect them against potential risk. For financial institutions, when their assets mainly consist of credit portfolios, they could manage to improve their profitability by keeping a reasonable amount of capital to satisfy various provisions, such as the coverage ratio, rather than having too much idle fund, to meet the risk supervision requirements. Because the marginal default probability and historical recovery rate of each asset in the portfolio are difficult to obtain, we make the following assumptions. First, we assume that the recovery rate is zero, meaning that all the risky assets are lost once the default occurs. Second, similar to Glasserman and Li (2005), we assume that the marginal default probability of the *i*-th firm is $p_i = 0.02 \cdot (1 + \sin(16\pi i/m))$, $i = 1, \ldots, 60$, which is between 0 and 0.04.

Since the standardized asset returns among the firms follow a multivariable *t*-distribution, through the Levenberg-Marquardt algorithm, we obtain the shifted mean vector $\mu = (\mu_1, \mu_2, \mu_3, \mu_4)$ shown in Table 3.

Table 3 Estimators of the elements of the shifted mean vector μ

Element	μ_1	μ_2	μ_3	μ_4
Estimator	0.0115	0.0032	0.00436	0.00237

Then using the proposed three-step IS algorithm for calculating credit portfolio risk based on the *t*-copula, we simulate the portfolio loss 4,000 times; at the same time, we conduct standard Monte Carlo simulation 10,000 times. We show in Table 4 and Figure 10 the obtained tail probabilities of the loss distribution $P(L > x)$, the corresponding threshold values of loss x, and the corresponding ES values for the credit portfolio.

Table 4 Estimating credit portfolio risk using the three-step IS procedure under the *t*-copula

$P(L > x)$	**Threshold value of loss x (billion)**	**ES (billion)**	**Variance ratios**
0.05	12567.3148	20286.8675	13.03
0.03	15273.4207	21229.4916	21.27
0.01	20009.1061	24431.3777	57.01
0.008	21362.1591	25467.6459	89.01
0.006	22038.6855	25933.1488	90.42
0.004	23391.7385	26858.4827	141.94

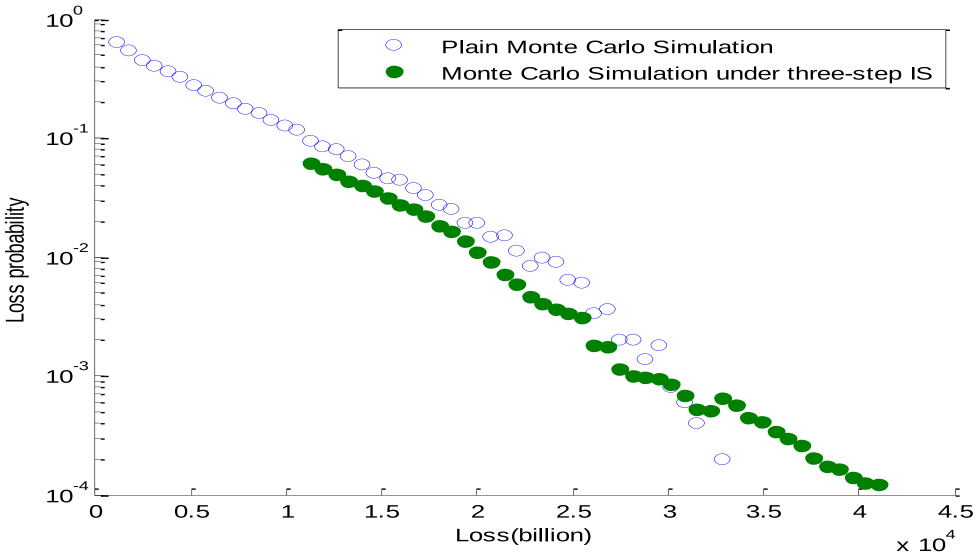

Figure 10 Tail probability of the loss distribution of the credit portfolio

The variance ratios in Table 4 show more than double-digit or triple-digit (one order of magnitude or two orders of magnitude) reductions in variance. This means that the three-step IS procedure achieves large variance reductions. Since the variance ratio is the ratio of the standard Monte Carlo variance to the IS variance, the larger the ratio is, the more effective is the proposed three-step IS procedure. So our proposed model is effective and efficient.

Accordingly, simulation precision is greatly improved. The key reason is that the rare events $\{L > x\}$ are more likely to occur under the new probability measure P_θ through exponential change of the probability measure, as explained by the theoretical analysis in Section 4.3.

Moreover, as shown in Table 4, the variance ratio increases with decreasing loss probability. This means that the smaller the loss probability is, the more effective is the three-step IS procedure. This coincides with the above theoretical analysis. As the loss probability becomes smaller, the corresponding loss value L of the portfolio increases, leading to reductions in the likelihood of the events $\{L > x\}$. Furthermore, as the loss probability decreases, the second-order moment with importance sampling decreases more rapidly than without importance sampling (i.e., by Eq. (12), the second-order moment with importance sampling decreases at twice the exponential rate of the loss probability itself). This causes the second-order moment $m_2(x,\theta)$ and the variance ratio in Table 4 to increase when the loss probability becomes smaller.

In addition, according to the definition of VaR, for a given loss probability, the threshold value of loss x is equivalent to VaR. The corresponding VaR and ES show that VaR is sensitive to changes in the loss probability, while ES does not fluctuate drastically with decreasing loss probability. This result again agrees with the theoretical analysis in Section 4.6.

According to the computational method of the portfolio ES in Section 6.2 and Eq. (29), we obtain the total provision for risk with consideration of credit risk and market risk in Table 5. As every day's default probability of each firm varies, the corresponding expected shortfalls and total provision for risks of portfolio are also changing. Based on the method in this paper, we can adjust the corresponding total provision for risk hedge against the risk in the supply chain every day.

Table 5 Estimating total provision for risks of portfolio

$P(L>x)$	**Total provision for risk (billion)**
0.05	16681.40279
0.03	19171.02021
0.01	23527.85078
0.008	24772.65954
0.006	25395.06383
0.004	26639.87259

Based on the ES value in Table 5, we calculate the total provision for risk by using Eq. (29). Thus, the obtained total provision for risk reacting in a timely manner to economic and environmental changes is to be obtained by synergizing the classical structural credit model, the integrating multivariate *t*-copula into the classical structural credit model, and the proposed three-step importance sampling (IS) model. Simultaneously, the eleven dimensions of data are being integrated together to calculate the joint default probability of credit portfolio and obtain the total provision for risk for adapting to economic and environmental changes. This also further illustrates that this data analytics technique is useful to extract and utilize useful information about credit portfolio risk by exploiting data from different sources. Namely, intelligent knowledge achieved through the data analytics is helpful to predict the risk level of the total credit portfolio. Our synergizing structural credit model can be widely applied in different situations and the corresponding results enable firms to better assess their own financial risk and those of their suppliers/customers, so that firms could make proper operational decisions in a timely manner, especially in facing rapid economic and environmental changes, and technology advancement.

8 Conclusions

In this paper we investigate how to improve operational risk management by giving a better estimation of the credit portfolio risk using data analytics. Specifically, we consider the decision maker that seeks to make the optimal operational decision on total provision for risk, based on his/her understanding of the credit portfolio risk. To do so, we investigate the impact of ES on operational decisions with total provision for risk. We build a structural credit model integrated with data analytics to estimate the tail probability of the total credit portfolio loss distribution. As such, we extend the two-step IS procedure into a three-step IS procedure for the *t*-copula portfolio credit risk measurement model in order to further reduce variance. Moreover, we apply the Levenberg-Marquardt algorithm, a nonlinear optimization technique, to find the optimal mean-shift vector of the systematic risk factors after the probability measure change to enhance the effectiveness of the IS procedure. We introduce the two risk measures VaR and ES to compute portfolio credit risk in the multivariate *t*-copula framework and compare their functionality. The results show that VaR is sensitive to changes in the loss probability, while ES does not fluctuate drastically with decreasing loss probability.

To assess the performance of our proposed model, we collect data from 60 listed Chinese firms and conduct simulation runs on the collected data using our model. Based on these real data, we calculate the tail probability of the credit portfolio loss distribution, and the corresponding VaR and ES. The simulation results show large variance reductions, which demonstrate that our three-step IS procedure based on the multivariable *t*-distribution is effective.

In addition, using the proposed *t*-copula portfolio credit risk measurement model, we check the changes in credit portfolio risk in different industries corresponding to the new national real estate regulation policies promulgated in China. The analytical results are in good agreement with the historical events that transpired, again illustrating that the proposed model is effective and reliable for practical use. With application of our model, firms could assess their financial risk better and calculate the total provision for risk with consideration of credit risk and market risk to prepare for risk loss reserve in a timely and adequate manner, so that they could make appropriate operational decisions when their financial status changes.

We calculate the credit portfolio risk of 60 listed Chinese firms in different industries using our proposed credit portfolio risk measure model in Empirical Tests section. One of the assumptions is that the recovery rate is zero, meaning that all the risky assets are lost once default occurs for the sake of calculation convenience. However, this may not reflect the real world.

In fact, the recovery rate is a key variable that determines the credit portfolio (Basel Committee on Banking Supervision, 2011; Chen et al., 2018). To predict changes in the recovery rate from historical defaulted loss, detecting the recovery rate density function is pivotal in credit risk analysis. Thus, in order to accurately calculate the tail probability of the credit portfolio loss distribution or make the corresponding optimal operational decisions with total provision for risk that hedges against the potential risk in the entire supply chain, one needs to integrate the time-variant and random recovery rates into our proposed credit portfolio risk measure model. Future research should consider how such integration should be done.

References

Azadegan, A., Parast, M.M., Lucianetti, L., Nishant, R., Blackhurst, J. (2020). Supply chain disruptions and business continuity: An empirical assessment. *Decision Sciences*, 51(1), 38-73.

Acerbi, C., Tasche, D. (2002) On the coherence of expected shortfall. *Journal of Banking & Finance, 26*(7), 1487-1503.

Aldrich, E.M., Heckenbach, I., Laughlin, G. (2016) A compound duration model for high-frequency asset returns. *Journal of Empirical Finance, 39,* 105-128.

Altman, E.I. (1968) Financial ratios, discriminant analysis and the prediction of corporate bankruptcy. *The Journal of Finance*, 23(4), 589-609.

Ankudinov, A., Ibragimov, R., Lebedev, O. (2017) Heavy tails and asymmetry of returns in the Russian stock market. *Emerging Markets Review*, 32, 200-219.

Basel Committee on Banking Supervision (2011). Basel III counterparty credit risk frequently asked questions. *Basel: Bank for International Settlements*, November 2011, Basel.

Bassamboo, A., Juneja, S., Zeevi, A. (2008) Portfolio credit risk with extremal dependence: Asymptotic analysis and efficient simulation. *Operations Research*, 56(3), 593-606.

Behr, A., Pötter, U. (2009) Alternatives to the normal model of stock returns: Gaussian mixture, generalized log F and generalised hyperbolic models. *Annals of Finance*, 5(1), 49-68.

Biglova A., Ortobelli, S., Rachev, S., Fabozzi, F. (2009) Modeling, estimation, and optimization of equity portfolios with heavy-tailed distributions. Chapter 5 in *Optimizing Optimization: The Next Generation of Optimization Applications and Theory*. Edited by Stephen Satchell, pp.117-141.

Black, F., Scholes, M. (1973) The pricing of options and corporate liabilities. *Journal of Political Economy*, 81(3), 637-654.

Boudreault, M., Gauthier, G., Thomassin, T. (2015) Estimation of correlations in portfolio credit risk models based on noisy security prices. *Journal of Economic Dynamics and Control*, 61(C), 334-349.

Broda, S.A. (2012) The expected shortfall of quadratic portfolios with heavy-tailed risk factors. *Mathematical Finance*, 22(4), 710-728.

Brusset, X., Bertrand, J.L. (2018) Hedging weather risk and coordinating supply chains. *Journal of Operations Management*, 64, 41-52.

Bülbül, D., Hakenes, H., Lambert, C. (2019) What influences banks' choice of credit risk management practices? Theory and evidence. *Journal of Financial Stability*, 40, 1-14.

Chan, J.C.C., Kroese, D.P. (2010) Efficient estimation of large portfolio loss probabilities in *t*-copula models. *European Journal Operational Research*, 205(2), 361-367.

Chen, R., Yang, L., Wang, W., Tang, L. (2015) Discovering the impact of systemic and idiosyncratic risk factors on credit spread of corporate bond within the framework of intelligent knowledge management. *Annals of Operations Research*, 234(1), 3-15.

Chen, R., Zhou, H., Jin, C., Zheng, W. (2018) Modeling of recovery rate for a given default by non-parametric method. *Pacific-Basin Finance Journal*, Available online, forthcoming.

Choi, T., Chan, H.K., Yue, X. (2017) Recent development in big data analytics for business operations and risk management. *IEEE Transactions on Cybernetics*, 47(1), 81-92.

D'Amico, G., Petroni, F. (2018) Copula based multivariate semi-Markov models with applications in high-frequency finance. *European Journal of Operational Research*, 267(2), 765-777.

Fernández, A., Rumí, R., Salmeró, A. (2012) Answering queries in hybrid Bayesian networks using importance sampling. *Decision Support Systems*, 53(3), 580-590.

Frey, R., McNeil, A.J. (2002) VaR and expected shortfall in portfolios of dependent credit risks: Conceptual and practical insights. *Journal of Banking and Finance*, 26(7), 1317-1334.

Glasserman, P., Heidelberger, P., Shahabuddin, P. (2000) Variance reduction techniques for estimating value-at-risk. *Management Science*, 46(10), 1349-1364.

Glasserman, P., Heidelberger, P., Shahabuddin, P. (2002) Portfolio value-at-risk with heavy-tailed risk factors. *Mathematical Finance*, 12(3), 239-269.

Glasserman, P., Kang, W., Shahabuddin, P. (2008) Fast simulation of multifactor portfolio credit risk.

Operational Research, 56(5), 1200-1217.

Glasserman P., Li, J.Y. (2005) Importance sampling for portfolio. *Management Science*, 51(11), 1643-1656.

Grundke, P. (2007) Computational aspects of integrated market and credit portfolio models. *OR Spectrum*, 29(2), 259-294.

Grundke, P. (2009) Importance sampling for integrated market and credit portfolio models. *European Journal of Operational Research*, 194(1), 206-226.

Goel, P., Datta, A., Mannan, M.S., (2017) Application of big data analytics in process safety and risk management. *Proceedings of 2017 IEEE International Conference on Big Data (BIGDATA)*, 1143-1152.

Hailemariam, G., Hill, S., Demissie, S. (2012) Exploring data mining techniques and algorithms for predicting customer loyalty and loan default risk scenarios at wisdom microfinance, Addis Ababa, Ethiopia. *MEDES '12 Proceedings of the International Conference on Management of Emergent Digital EcoSystems*, Addis Ababa, Ethiopia, 183-184.

Hillegeist, S.A, Keating, E.K., Cram, D.P., Lundstedt, K.G. (2004) Assessing the probability of bankruptcy. *Review of Accounting Studies*, 9(1), 5–34.

Hsieh, M. H., Lee, Y. H., Shyu, S. D., Chiu, Y. F. (2018) Estimating multifactor portfolio credit risk: A variance reduction approach. *Pacific-Basin Finance Journal*, in press, DOI: 10.1016/j.pacfin.2018.08.001.

Hu, Q. (2016) Project portfolio selection: A newsvendor approach. *Decision Sciences*, 48(1), 176-199.

Huang, Y., Kou, G. (2014) A kernel entropy manifold learning approach for financial data analysis.

Decision Support Systems, 64, 31-42.

Huang, Z., Shahabuddin, P. (2003) New Simulation methodology for risk analysis: Rare-event, heavy-tailed simulations using hazard function transformations, with applications to Value-at-risk. *Proceedings of the 35th Conference on Winter Simulation: Driving Innovation. New Orleans: Winter Simulation Conference*, 276-284.

Ivanov, D., Dolgui, A., Sokolov, B. (2018). The impact of digital technology and Industry 4.0 on the ripple effect and supply chain risk analytics. *International Journal of Production Research*, in press, DOI:10.1080/00207543.2018.1488086.

Johannes, V.S., Jeffrey, T.L., Anna, S.-H. (2009) Value-at-Risk computation by Fourier inversion with explicit error bonds. *Finance Research Letters*, 6(2), 95-105.

Jones, E.P., Mason, S.P., Rosenfeld, E. (1984) Contingent claims analysis of corporate capital structures: An empirical investigation. *Journal of Finance*, 39(3), 611-625.

Jules, S.K. (2012) VaR and ES for linear portfolios with mixture of generalized Laplace distributions risk factors. *Annals of Finance*, 8 (1), 123–150.

Kalkbrener, M., Lotter, H., Overbeck, L. (2004) Sensible and efficient capital allocation for credit portfolios. *Journal of Risk*, 17(1), 19-24.

Kamdem, J.S. (2009) Δ-VaR and Δ-TVaR for portfolios with mixture of elliptic distributions risk factors and DCC. *Insurance: Mathematics and Economics*, 44(3), 325-336.

Kang, W., Shahabuddin, P. (2005) Fast simulation for multifactor portfolio credit risk in the *t*-copula model. *Proceedings of the 37th Conference on Winter Simulation*, December 04-07, Orlando, 1859-1867.

Kou, G., Chao, X., Peng, Y., Alsaadi, F. E., Herrera-Viedma, E. (2019) Machine learning methods for systemic risk analysis in financial sectors. *Technological and Economic Development of Economy*, 25(5), 716-742.

Kouvelis, P., Pang, Z., Ding, Q. (2018) Integrated commodity inventory management and financial hedging: A dynamic mean-variance analysis. *Production and Operations Management*, 27(6), 1052-1073.

Koyuncugil, A.S., Ozgulbas, N. (2012) Financial early warning system model and data mining application for risk detection. *Expert Systems with Applications*, 39(6), 6238-6253.

Kole, E., Koedijk, K., Verbeek, M. (2007) Selecting copulas for risk management. *Journal of Banking & Finance*, 31(8), 2405-2423.

Kostadinov, K. (2006) Tail approximation for credit risk portfolios with heavy-tailed risk factors. *Journal of Risk*, 8(2), 187-210.

Kumar, R.L., Park, S. (2019) A portfolio approach to supply chain risk management, *Decision Sciences*, in press, DOI: 10.111/deci.12332.

Leveling, J., Edelbrock, M., Otto, B. (2014) Big data analytics for supply chain management. *Proceedings of the 2014 IEEE IEEM*, 918-922.

Levenberg, K. (1944) A method for the solution of certain problems in least squares. *Quarterly of Applied Mathematics*, 2, 164-168.

Li, L., Ren, F. (2018) Costly operations burden bike-sharing firm Ofo, *National Business Daily*. Retrieved 19 Feb 2019.

Ma, S., Li, G., Sethi, S.P., Zhao, X. (2019) Advance selling in the presence of market power and risk-averse consumers. *Decision Sciences*, 50(1), 142-169.

Marquardt, D.W. (1963) An algorithm for least-squares estimation of nonlinear parameters. *Journal of the Society for Industrial & Applied Mathematics*, 11(2), 431-441.

McNamara, J.R. (1998) Portfolio selection using stochastic dominance criteria. *Decision Sciences*, 29(4), 785-801.

Merino, S., Nyfeler, M. (2002) Calculating portfolio loss. *Journal of Risk*, 15(8), 82-86.

Merton, R.C. (1974) On the pricing of corporate debt: the risk structure of interest rates. *Journal of Finance*, 29, 449-470.

Ohlson, J.A. (1980) Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research*, 18, 109-131.

Osmundsen, K.K. (2018) Using expected shortfall for credit risk regulation. *Journal of International Financial Markets*, Institutions and Money, 57, 80-93.

Reitan, T., Aas, K. (2010) A new robust importance-sampling method for measuring value-at-risk and expected shortfall allocations for credit portfolios. *Journal of Credit Risk*, 6(4), 113-149.

Rosenberg, J.V., Schuermann, T.A. (2006) General approach to integrated risk management with skewed, fat-tailed risks. *Journal of Financial Economics*, 79(3), 569-614.

Schneider, J.C., Schweizer, N. (2015) Robust measurement of (heavy-tailed) risks: Theory and implementation. *Journal of Economic Dynamics & Control*, 61(12), 183-203.

Smith, L.D., Sanchez, S.M., Lawrence, E.C. (1996) A comprehensive model for managing credit risk on home mortgage portfolios. *Decision Sciences*, 27(2), 291-317.

Sun, L., Zheng, X., Jin, Y., Jiang, M., Wang, H. (2019) Estimating promotion effects using big data: A partially profiled LASSO model with endogeneity correction. *Decision Sciences*, in press, DOI: 10.1111/deci.12354.

Sun, X., Chung, S., Ma, H., (2020) Operational risk in airline crew scheduling: Do features of flight delays matter? *Decision Sciences*, in press, DOI: 10.1111/deci.12426.

Vassalou, M., Xing, Y. (2004) Default risk in equity returns. *The Journal of Finance*, 59(2), 831-868.

Wang, L., Yao, D.D., (2019). Risk hedging for production planning. *Production and Operations Management*, in press, DOI: 10.1111/poms.13103.

Wang, T., Wu, D. (2020). Data-driven driver dispatching system with allocation constraints and operational risk management for a ride-sharing platform. *Decision Science*, in press, DOI: 10.1111/deci.12433.

Xie, F., He, Z, Wang, X. (2019) An importance sampling-based smoothing approach for quasi-Monte Carlo simulation of discrete barrier options. *European Journal of Operational Research*, 274(2), 759-772.

Appendix 1 Proof of Eq. (20)

$$f_\alpha(V) = f(V)\exp(\alpha V - \psi_V(\alpha)) = \frac{V^{(\nu/2-1)}\exp(-V/2)}{2^{\nu/2}\Gamma(\nu/2)}\exp(\alpha V - \psi_V(\alpha))$$

$$= \frac{V^{(\nu/2-1)}\exp(-(\frac{1-2\alpha}{2})V)}{2^{\nu/2}\Gamma(\nu/2)}\exp(-\psi_V(\alpha)) = \frac{V^{(\nu/2-1)}\exp(-(\frac{1-2\alpha}{2})V)}{2^{\nu/2}\Gamma(\nu/2)}\frac{1}{\phi_V(\alpha)}$$

$$= \frac{V^{(\nu/2-1)}\exp(-(\frac{1-2\alpha}{2})V)}{2^{\nu/2}\Gamma(\nu/2)}(1-2\alpha)^{\nu/2} = \frac{V^{(\nu/2-1)}\exp(-(\frac{1-2\alpha}{2})V)}{\Gamma(\nu/2)}(\frac{1-2\alpha}{2})^{\nu/2}$$

$$= (\frac{1-2\alpha}{2})^{\nu/2}\frac{V^{(\nu/2-1)}}{\Gamma(\nu/2)}\exp(-(\frac{1-2\alpha}{2})V)$$

$$= (\frac{2}{1-2\alpha})^{-\nu/2}\frac{V^{(\nu/2-1)}}{\Gamma(\nu/2)}\exp(-\frac{V}{2/(1-2\alpha)}).$$

This completes the proof.

Appendix 2 60 Listed firms

Details of the 60 listed Chinese firms are given in the following table. The 60 firms, listed in the Shanghai Stock Exchange, include 25 firms from the real estate industry (serial numbers 1 to 25), 20 firms from the retail industry (serial numbers 26 to 45), and 15 firms from the finance and insurance industry (serial numbers 46 to 60).

Serial number	Listed firm name	Industry	Serial number	Listed firm name	Industry
1	Poly Real Estate	real estate	31	Tailong Pharmaceutical	retail
2	Zhejiang Guangsha	real estate	32	Nanjing Zhongshang	retail
3	Songdu shares	real estate	33	Jiangsu Haotian	retail
4	Daming City	real estate	34	Commercial city	retail
5	Xiangjiang Holdings	real estate	35	Great Eastern	retail
6	Xinhu Zhongbao	real estate	36	Meike shares	retail
7	Lushang Real Estate	real estate	37	Hualian Comprehensive	retail
8	Wantong Real Estate	real estate	38	Guangdong Pearl	retail
9	Beijing Urban Construction	real estate	39	Sinopharm	retail
10	Huafa Shares	real estate	40	New world	retail
11	Gemdale Group	real estate	41	Nanjing Xinbai	retail
12	Qixia Construction	real estate	42	Dongbai Group	retail
13	Dingli	real estate	43	Eurasian Group	retail
14	Fenghua shares	real estate	44	Property in the big	retail
15	Wanye Enterprise	real estate	45	Nanning Department Store	retail

16	Cinda Real Estate	real estate	46	Shanghai Pudong Development Bank	finance and insurance
17	Heaven and earth source	real estate	47	Minsheng Bank	finance and insurance
18	Chinese company	real estate	48	CITIC Securities	finance and insurance
19	Zhujiang Industry	real estate	49	China Merchants Bank	finance and insurance
20	Doron shares	real estate	50	Guojin Securities	finance and insurance
21	Phoenix shares	real estate	51	Southwest Securities	finance and insurance
22	Shanghai Xinmei	real estate	52	Haitong Securities	finance and insurance
23	Huayuan Real Estate	real estate	53	China Merchants Securities	finance and insurance
24	Tibet City Investment	real estate	54	Bank of Nanjing	finance and insurance
25	Shimao shares	real estate	55	Agricultural Bank of China	finance and insurance
26	Minmetals Development	retail	56	Ping An	finance and insurance
27	Zhejiang Oriental	retail	57	Bank of Communications	finance and insurance
28	Hongtu Hi-Tech	retail	58	ICBC	finance and insurance
29	Hongye shares	retail	59	Bank of China	finance and insurance
30	Jianfa shares	retail	60	CITIC Bank	finance and insurance

1

2

3

4

5

6

Appendix 3 Normality test

	Statistics		KS-test		JB-test		Lillie-test	
Firm name	Skewness	Kurtosis	h	P	h	P	h	p
Poly Real Estate	-0.83	6.8	1	0.000678	1	0.001	1	0.001
Zhejiang Guangsha	1.67	8.1	1	0.000928	1	0.001	1	0.001
Songdu shares	0.22	3.1	1	0.000608	1	0.001	1	0.001
Daming City	1.89	9.8	1	0.000644	1	0.001	1	0.001
Xiangjiang Holdings	2.15	10.9	1	0.000669	1	0.001	1	0.001
Xinhu Zhongbao	0.13	4.8	1	0.000785	1	0.001	1	0.001
Lushang Real Estate	-1.66	14.3	1	0.000695	1	0.001	1	0.001
Wantong Real Estate	-0.02	6.0	1	0.000603	1	0.001	1	0.001
Beijing Urban Construction	0.38	3.2	1	0.000459	1	0.001	1	0.001
Huafa Shares	1.45	6.6	1	0.000964	1	0.001	1	0.001
Gemdale Group	-1.46	6.0	1	0.000634	1	0.001	1	0.001
Qixia Construction	3.28	20.1	1	0.001238	1	0.001	1	0.001
Dingli	0.15	4.3	1	0.001163	1	0.001	1	0.001
Fenghua shares	0.02	4.3	1	0.000812	1	0.001	1	0.001
Wanye Enterprise	-0.05	6.3	1	0.001139	1	0.001	1	0.001
Cinda Real Estate	1.35	5.5	1	0.000592	1	0.001	1	0.001
Heaven and earth source	0.25	3.5	1	4.76E-05	1	0.001	1	0.001
Chinese company	0.93	9.1	1	0.000902	1	0.001	1	0.001
Zhujiang Industry	1.53	8.0	1	0.000334	1	0.001	1	0.001
Doron shares	0.41	4.4	1	0.000703	1	0.001	1	0.001
Phoenix shares	1.79	6.9	1	0.000228	1	0.001	1	0.001
Shanghai Xinmei	0.58	4.5	1	0.001249	1	0.001	1	0.001
Huayuan Real Estate	-0.07	4.3	1	0.001143	1	0.001	1	0.001
Tibet City Investment	0.07	6.6	1	0.000725	1	0.001	1	0.001
Shimao shares	0.55	3.3	1	0.00144	1	0.001	1	0.001
Minmetals Development	0.21	3.6	1	0.000139	1	0.001	1	0.001
Zhejiang Oriental	0.38	4.5	1	0.000351	1	0.001	1	0.001

Hongtu Hi-Tech	1.14	5.4	1	0.000798	1	0.001	1	0.001
Hongye shares	-0.32	5.5	1	0.000257	1	0.001	1	0.001
Jianfa shares	-0.70	6.7	1	0.000226	1	0.001	1	0.001
Tailong Pharmaceutical	0.64	5.2	1	0.000832	1	0.001	1	0.001
Nanjing Zhongshang	-0.97	5.8	1	0.000929	1	0.001	1	0.001
Jiangsu Haotian	0.51	4.7	1	0.000547	1	0.001	1	0.001
Commercial city	2.19	9.3	1	0.000562	1	0.001	1	0.001
Great Eastern	0.75	5.5	1	0.000396	1	0.001	1	0.001
Meike shares	0.13	6.9	1	0.001252	1	0.001	1	0.001
Hualian Comprehensive	1.35	5.8	1	0.000874	1	0.001	1	0.001
Guangdong Pearl	-0.61	5.0	1	0.001216	1	0.001	1	0.001
Sinopharm	0.30	3.5	1	0.000258	1	0.001	1	0.001
New world	0.21	5.1	1	0.001331	1	0.001	1	0.001
Nanjing Xinbai	2.74	13.8	1	0.000892	1	0.001	1	0.001
Dongbai Group	-1.02	5.6	1	0.000629	1	0.001	1	0.001
Eurasian Group	-0.68	9.5	1	0.00038	1	0.001	1	0.001
Property in the big	-1.36	9.3	1	0.000795	1	0.001	1	0.001
Nanning Department Store	0.20	4.7	1	0.000762	1	0.001	1	0.001
Shanghai Pudong Development Bank	2.18	10.8	1	0.00037	1	0.001	1	0.001
Minsheng Bank	1.55	6.2	1	0.000501	1	0.001	1	0.001
CITIC Securities	-0.77	8.5	1	0.000909	1	0.001	1	0.001
China Merchants Bank	-0.35	5.2	1	0.001028	1	0.001	1	0.001
Guojin Securities	1.57	7.7	1	0.00128	1	0.001	1	0.001
Southwest Securities	-0.19	3.5	1	0.000563	1	0.001	1	0.001
Haitong Securities	-1.52	8.3	1	0.001258	1	0.001	1	0.001
China Merchants Securities	-0.51	9.4	1	0.001022	1	0.001	1	0.001
Bank of Nanjing	1.41	8.6	1	0.000651	1	0.001	1	0.001
Agricultural Bank of China	2.54	14.0	1	0.001818	1	0.001	1	0.001

Ping An	-1.23	12.9	1	0.000583	1	0.001	1	0.001
Bank of Communications	0.09	4.0	1	0.001359	1	0.001	1	0.001
ICBC	0.16	4.2	1	0.001214	1	0.001	1	0.001
Bank of China	-0.15	3.7	1	0.001279	1	0.001	1	0.001
CITIC Bank	1.53	7.9	1	0.000266	1	0.001	1	0.001

Appendix 4 Parameter estimation of copula function

	Systematic risk factors				Factor loading
Firm name	a_{i1}	a_{i2}	a_{i3}	a_{i4}	b_i
Poly Real Estate	0.026	0.003	0.001	0.000	0.9997
Zhejiang Guangsha	0.002	0.000	0.011	0.154	0.9879
Songdu shares	0.090	0.000	0.031	0.012	0.9954
Daming City	0.014	0.000	0.028	0.000	0.9995
Xiangjiang Holdings	0.044	0.000	0.000	0.000	0.9990
Xinhu Zhongbao	0.000	0.045	0.007	0.143	0.9887
Lushang Real Estate	0.027	0.000	0.000	0.002	0.9996
Wantong Real Estate	0.000	0.160	0.115	0.281	0.9393
Beijing Urban Construction	0.058	0.000	0.000	0.007	0.9983
Huafa Shares	0.032	0.000	0.000	0.000	0.9995
Gemdale Group	0.489	0.000	0.003	0.147	0.8599
Qixia Construction	0.000	0.001	0.000	0.289	0.9575
Dingli	0.000	0.030	0.031	0.004	0.9991
Fenghua shares	0.000	0.138	0.185	0.400	0.8871
Wanye Enterprise	0.000	0.165	0.181	0.420	0.8740
Cinda Real Estate	0.000	0.122	0.213	0.149	0.9578
Heaven and earth source	0.000	0.122	0.222	0.357	0.8989
Chinese company	0.000	0.153	0.152	0.240	0.9464
Zhujiang Industry	0.000	0.002	0.000	0.035	0.9994
Doron shares	0.000	0.128	0.119	0.210	0.9621
Phoenix shares	0.000	0.252	0.347	0.438	0.7898
Shanghai Xinmei	0.000	0.007	0.000	0.067	0.9977
Huayuan Real Estate	0.001	0.001	0.000	0.032	0.9995
Tibet City Investment	0.000	0.003	0.000	0.432	0.9020
Shimao shares	0.000	0.087	0.102	0.231	0.9638
Minmetals Development	0.002	0.000	0.000	0.011	0.9999
Zhejiang Oriental	0.000	0.001	0.096	0.176	0.9797
Hongtu Hi-Tech	0.000	0.000	0.015	0.150	0.9885
Hongye shares	0.000	0.052	0.358	0.666	0.6525
Jianfa shares	0.000	0.008	0.001	0.014	0.9999

Tailong Pharmaceutical	0.000	0.164	0.150	0.398	0.8903
Nanjing Zhongshang	0.035	0.010	0.067	0.000	0.9971
Jiangsu Haotian	0.000	0.059	0.055	0.254	0.9638
Commercial city	0.684	0.013	0.000	0.245	0.6870
Great Eastern	0.000	0.000	0.178	0.444	0.8784
Meike shares	0.000	0.044	0.048	0.161	0.9848
Hualian Comprehensive	0.000	0.129	0.156	0.152	0.9674
Guangdong Pearl	0.086	0.000	0.015	0.000	0.9962
Sinopharm	0.000	0.119	0.254	0.230	0.9320
New world	0.000	0.159	0.195	0.447	0.8583
Nanjing Xinbai	0.000	0.031	0.160	0.307	0.9378
Dongbai Group	0.036	0.000	0.000	0.055	0.9978
Eurasian Group	0.304	0.000	0.000	0.038	0.9518
Property in the big	0.000	0.000	0.000	0.192	0.9814
Nanning Department Store	0.000	0.123	0.328	0.422	0.8360
Shanghai Pudong Development Bank	0.000	0.000	0.000	0.010	0.9999
Minsheng Bank	0.000	0.016	0.000	0.033	0.9993
CITIC Securities	0.000	0.000	0.000	0.001	1.0000
China Merchants Bank	0.000	0.017	0.012	0.018	0.9996
Guojin Securities	0.000	0.001	0.002	0.233	0.9724
Southwest Securities	0.000	0.023	0.000	0.016	0.9996
Haitong Securities	0.068	0.006	0.000	0.000	0.9977
China Merchants Securities	0.000	0.000	0.029	0.000	0.9996
Bank of Nanjing	0.013	0.000	0.000	0.025	0.9996
Agricultural Bank of China	0.000	0.045	0.001	0.078	0.9959
Ping An	0.007	0.017	0.000	0.023	0.9996
Bank of Communications	0.000	0.004	0.000	0.000	1.0000
ICBC	0.000	0.000	0.000	0.003	1.0000
Bank of China	0.001	0.113	0.000	0.087	0.9898
CITIC Bank	0.052	0.003	0.027	0.000	0.9983

1