

Article

Predicting China's SME Credit Risk in Supply Chain Financing by Logistic Regression, Artificial Neural Network and Hybrid Models

You Zhu ¹, Chi Xie ^{1,2,*}, Bo Sun ^{3,4}, Gang-Jin Wang ^{1,2} and Xin-Guo Yan ¹

¹ College of Business Administration, Hunan University, Changsha 410082, China; zy19@hnu.edu.cn (Y.Z.); wanggangjin@hnu.edu.cn (G.-J.W.); yandu@hnu.edu.cn (X.-G.Y.)

² Center of Finance and Investment Management, Hunan University, Changsha 410082, China

³ Economics and Management School, Wuhan University, Wuhan 430072, China; bsun0916@yahoo.com

⁴ China Huarong Asset Management CO., LTD., Beijing 100033, China

* Correspondence: xiechi@hnu.edu.cn; Tel.: +86-731-8882-3890

Academic Editor: Giuseppe Ioppolo

Received: 18 February 2016; Accepted: 26 April 2016; Published: 3 May 2016

Abstract: Based on logistic regression (LR) and artificial neural network (ANN) methods, we construct an LR model, an ANN model and three types of a two-stage hybrid model. The two-stage hybrid model is integrated by the LR and ANN approaches. We predict the credit risk of China's small and medium-sized enterprises (SMEs) for financial institutions (FIs) in the supply chain financing (SCF) by applying the above models. In the empirical analysis, the quarterly financial and non-financial data of 77 listed SMEs and 11 listed core enterprises (CEs) in the period of 2012–2013 are chosen as the samples. The empirical results show that: (i) the “negative signal” prediction accuracy ratio of the ANN model is better than that of LR model; (ii) the two-stage hybrid model type I has a better performance of predicting “positive signals” than that of the ANN model; (iii) the two-stage hybrid model type II has a stronger ability both in aspects of predicting “positive signals” and “negative signals” than that of the two-stage hybrid model type I; and (iv) “negative signal” predictive power of the two-stage hybrid model type III is stronger than that of the two-stage hybrid model type II. In summary, the two-stage hybrid model III has the best classification capability to forecast SMEs credit risk in SCF, which can be a useful prediction tool for China's FIs.

Keywords: supply chain financing (SCF); credit risk; small and medium-sized enterprises (SMEs); core enterprises (CEs); financial institutions (FIs); logistic regression (LR); artificial neural network (ANN); two-stage hybrid model

1. Introduction

Intense market competition, capital shortages and globalization generate complex and dynamic supply chains. Reinforcing the management of material flow and information flow does not necessarily result in improving the management of supply chain. Therefore, the focus of supply chain management today is on the design and optimization of cash flow [1]. Supply chain financing (SCF) has increasingly become a hot topic in supply chain management and a growing product category of financial institutions (FIs). In China, SCF is experiencing a rapid development stage and numerous FIs have begun to focus on developing and designing new SCF services and products to solve the financing issues facing SMEs (e.g., 1 + N SCF of the Pingan Bank). SCF is a type of channel for financing, which manages, plans and controls all cash flows across supply chain members to improve the turnover efficiency of working capital [2]. In SCF, small and medium-sized enterprises (SMEs) obtain loans with looser constraints from banks through expanded credit lines, core enterprises (CEs) alleviate the pressure of funding, and financial intermediaries dramatically increase their incomes [3–5]. More

specifically, SCF significantly decreases the credit risk of SMEs for FIs [6]. Nevertheless, SCF cannot completely eliminate credit risks, which continue to be one of the major threats to FIs [7–9]. Moreover, SCF has been promoted for almost ten years and has experienced slow development in China because we do not have an appropriate SME credit risk evaluation index system or an outstanding prediction model, which hinder SCF.

The Basel Committee on Banking Supervision Principles for the Management of Credit Risk defines credit risk as the possibility that the borrower or the lender will fail to keep an appointment with the correlative bank. In China, SMEs are the main applicants of SCF, thus the bank suffers from credit risk in SCF when the SMEs cannot honor an agreement. Researchers and bankers emphasize that structuring the SME credit risk evaluation index system is the largest and most critical challenge to banks' management of SCF and is the fundamental work in credit loan decision making. A good credit risk evaluation index system can guarantee profitability and stability of a FI, whereas a poor system can potentially lead to losses [10–12]. We propose that the SME credit risk evaluation index system of SCF evaluate credit risks from various aspects, including the SMEs' financial condition, the CEs' financial condition, the operational status of the entire supply chain and the transactional relationship between the SMEs and the CEs.

In the field of structuring SME credit risk evaluation index systems, numerous studies focus on applying or integrating data mining tools to improve the SME credit risk prediction accuracy ratio of existing models. The credit risk prediction accuracy ratio is a ratio that predicts dichotomous outcomes of good and bad credit cases in a financing market. The credit risk prediction accuracy ratio is calculated based on a cumulative accuracy profile (CAP) curve, which is constructed by sorting the debtors in order from bad credit classes to good credit classes, *i.e.*, by decreasing credit risk [13]. Logistic regression (LR) is widely used because it is an efficient and robust method of prediction [14]. Many studies have focused on analyzing the default probability estimation of SMEs using the LR method, which provide available credit risk prediction in finance and involve different countries' cases. For example, in 2005, Altman and Sabato [15] used data from the USA, Italy and Australia to investigate the effects of the Basel II on bank capital requirements for SMEs using the LR method; in 2007, Altman and Sabato [16] proposed a new distress prediction model specifically for SMEs; and Bebr and Güttler [17] applied a logistic scoring model for predicting the probability of default using a data set for German SMEs. In addition, Fantazzini and Figini [18], Fidrmuc and Heinz [19], Pederzoli and Torricelli [20], Pederzoli and Thoma [21] also analyzed the default probability of SMEs using the LR method. However, the credit risk prediction accuracy ratio of the LR approach is lower than that of the artificial neural network (ANN) method [22]. The ANN method is also widely applied in the default prediction of SMEs. For example, early on, Salchenberger *et al.* [23] applied an ANN for predicting thrift failures and supporting SMEs to make a correct decision. Sharda and Wilson [24] analyzed predictive performance measurement issues and conducted ANN experiments in business failure forecasting. Zhang *et al.* [25] applied the ANN method to bankruptcy prediction and, early on, applied cross-validation analysis. Although ANN methods provide a strong credit risk prediction capability, they are criticized for their long training process when designing the optimal network topology, which limits its applicability in handling credit risk prediction problems [26–28]. The respective characteristics of LR and ANN lead scholars to combine these two methods to measure credit risk for FIs. For instance, Lin [29] and Falavigna [30] proposed two-stage hybrid models by combining LR and ANN approaches and explored whether the two-stage hybrid model outperforms traditional LR and ANN methods. Researchers usually use the LR model or the ANN model to forecast China's SME credit risk in SCF. For example, Deng *et al.* [31] and Bai and Li [32] considered that the ANN model is suitable for predicting SME credit risk at the present stage for China's credit market. Nevertheless, Xiong *et al.* [33], Bai [34] and Bei *et al.* [35] proposed the LR model instead of an ANN model and argued that the robustness is more important than the accuracy of the model for the early stage of SCF. Unfortunately, we did not find any studies that applied the two-stage hybrid model in predicting China's SME credit risk in SCF. In our study, we explore the

SME credit risk prediction performance of the LR model, the ANN model and three types of two-stage hybrid models for China's FIs in SCF. Our prediction is based on the estimation of the condition mean. Other prediction methods, for example, a method based on the conditional quantile estimation, are possible [36,37].

The contributions of this paper are summarized as follows: (1) we propose an SME credit risk evaluation index system specifically for SCF. This system is used to evaluate the credit risks from different points of view, which not only consist of SMEs' financial and non-financial conditions but also contain CEs' financial and non-financial conditions, the operational status of the entire supply chain, and the transactional relationship between SMEs and CEs; (2) we demonstrate that the SME credit risk prediction performance of the type-III two-stage hybrid model is also better than that of the LR and ANN models and that of the type-I and type-II two-stage hybrid models in SCF.

The results of this paper show the following: (1) SCF has the ability to reduce the credit risk of SMEs, but it cannot completely eliminate credit risk, which remains a threat to FIs and the entire supply chain; (2) the primary empirical results show that the finance decision made by FIs mainly depends on the financial and non-financial conditions of the CEs in SCF; (3) and the two-stage hybrid model can provide a new perspective for improving the prediction accuracy ratio of China's SME credit risk in SCF. Overall, in practical terms, the two-stage hybrid model can be applied in credit risk prediction in SCF and is advantageous to addressing the trouble of the slow promotion of SCF in China.

The remainder of the paper is organized as follows. Section 2 discusses the methodology. Section 3 presents the description of the data and sampling procedure. The empirical results are shown in Section 4. Finally, Section 5 draws some conclusions.

2. Methodology

2.1. Logistic Regression (LR) Model

Regression methods attempt to describe the relationship between a response variable and one or more explanatory variables [38]; LR is a prevalent regression model. LR can be generally classified into binary logistic regression and multinomial logistic regression depending on whether dependent variables are dichotomous or multinomial [39]. In this study, dependent variables are dichotomous: the "negative signal" class takes on a value of 0, and the "positive signal" class takes on a value of 1. The LR model is represented as follows:

$$\ln(p/p') = \beta_0 + \sum_{j=1}^i c_j \beta_j \quad (1)$$

$$p = 1/[1 + e^{-(\beta_0 + \sum_{j=1}^i c_j \beta_j)}] \quad (2)$$

$$p' = 1 - p = 1/[1 + e^{(\beta_0 + \sum_{j=1}^i c_j \beta_j)}] \quad (3)$$

where p is the credit repayment compliance probability of SMEs; p' is the credit repayment default probability of SMEs; c_j ($j = 1, \dots, i$) is the j -th independent variable; β_0 is the intercept; β_j ($j = 1, \dots, i$) is the j -th coefficient associated with the j -th corresponding predictor c_j ($j = 1, \dots, i$); and $\ln(p/p')$ represents the credit risk signal. $\ln(p/p')$ being 0 denotes a "negative signal"; conversely, a value of 1 denotes a "positive signal".

Following Lin [29], we use the LR with the Wald-forward method to improve the performance of the LR model and select significant variables for constructing the subsequent two-stage hybrid models. The Wald-forward method is a stepwise selection procedure that continuously selects a single variable for the LR model in each step and uses the probability of the Wald statistic for selecting variables.

2.2. Artificial Neural Network (ANN) Model

As a class of beneficial non-linear modeling tools, ANN provides advantages when applied to prediction in a number of business areas and is capable of detecting all possible interactions between independent variables [40]. ANN has various network architectures such as multilayer perceptron (MLP) and radial basis function (RBF). In this paper, we test the RBF network for two reasons: first, the main disadvantage of the MLP network is that its local minima are limited and that its astringency is slow [41]; second, the RBF network performs better than the MLP network in terms of approximation capability, classification capacity and learning rate [42]. Broomhead and Lowe [43] initially applied the RBF network, whose neuron model and network structure are illustrated in Figure 1.

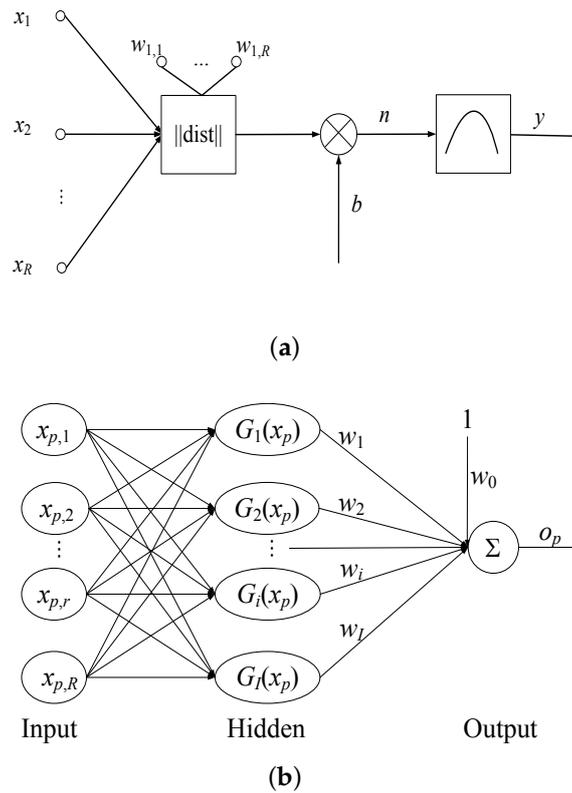


Figure 1. A neuron model of radial basis function (RBF) network is considered and includes a neuron with R input signals, the R input signals with w connection weights, the Euclidean distance between input vectors and weight vectors with $||dist||$, the threshold value as b , the independent variable of the activation function as n and the activation function as y . The structure of a three-layer radial basis neural network is considered and includes an input layer with x_p input variables, a single hidden layer with G_I Gaussian radial basis functions, output weights as w_I and an output layer with one neuron. (a) a neuron model of radial basis function network; (b) the structure of a three-layer radial basis neural network.

The RBF network’s activation function is usually a type of Gaussian radial basis function that can be represented as

$$G_i(||x_p - c_i||) = \exp\{-[1/(2\delta_i^2)]||x_p - c_i||^2\} \tag{4}$$

where G_i is the i -th Gaussian function, $||x_p - c_i||$ is the Euclidean norm, δ_i is the variance of the i -th Gaussian function, x_p is the p -th input sample, and c_i is the center or average of the i -th Gaussian RBF transformation.

A RBF network is composed of three layers: the input layer contains p input vectors, which have R input signals; the hidden layer is the radial basis layer, which contains I neurons with Gaussian functions; and the output layer is the linearity layer, which is a summing unit of the output weights w_l multiplied by the activation function outputs.

2.3. Two-Stage Hybrid Model

The two-stage hybrid model consists of two stages: in stage one, influencing variables are selected using LR with the Wald-forward method; in stage two, influencing variables are taken as the input variables of the ANN model (*i.e.*, the RBF) [22,26,29]. In this paper, we apply three types of two-stage hybrid models, namely model I, model II and model III. We illustrate these three types of models in Figure 2.

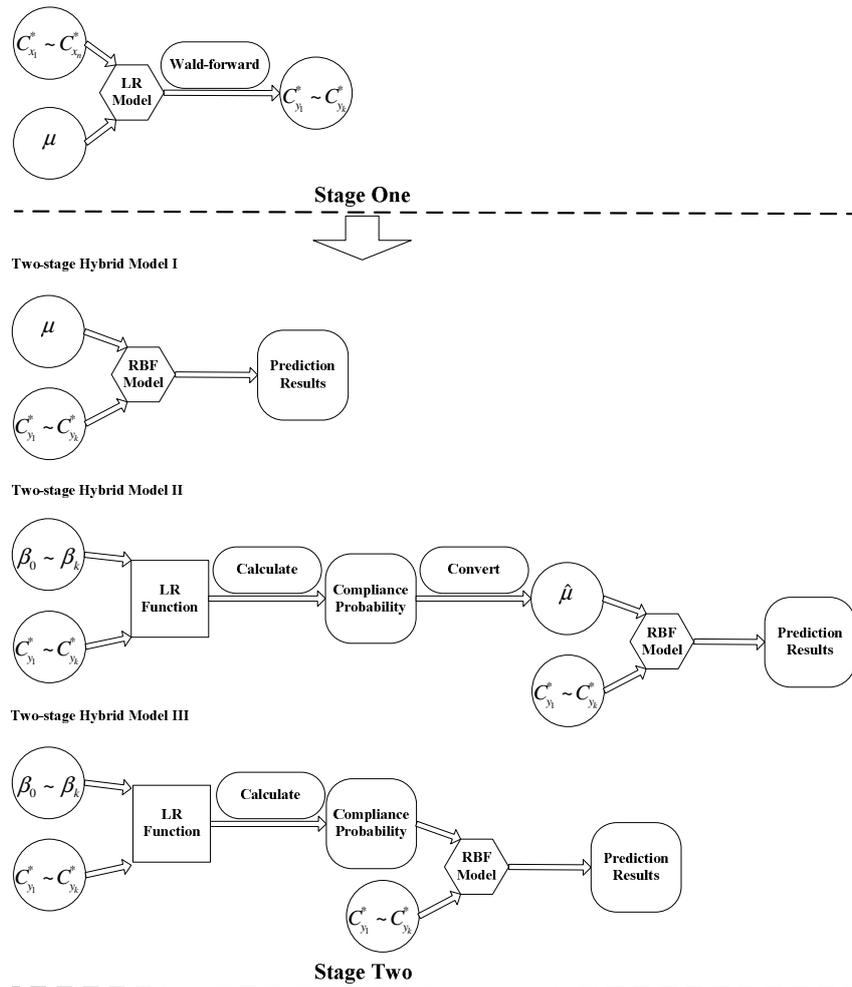


Figure 2. The two-stage hybrid models of LR-ANN (integrated by logistic regression and artificial neural network) are illustrated. In stage one, the independent variables $C_{x_1}^* \sim C_{x_n}^*$ and the dependent variable μ are substituted into the LR model using the Wald-forward method to identify the independent variables $C_{y_1}^* \sim C_{y_k}^*$ that significantly influence the compliance probability. In stage two, the two-stage hybrid model I is established using μ and $C_{y_1}^* \sim C_{y_k}^*$; the two-stage hybrid model II is established using $\hat{\mu}$ and $C_{y_1}^* \sim C_{y_k}^*$; the two-stage hybrid model III is established using the prediction value of compliance probability and $C_{y_1}^* \sim C_{y_k}^*$; the prediction value of compliance probability for each data sample is calculated by LR function; and the new dependent variable $\hat{\mu}$ is obtained by converting the compliance probability into a “negative signa” (value of 0) or a “ositive signa” (value of 1).

2.3.1. Two-Stage Hybrid Model of LR-ANN I

Model I is constructed using the following procedure: (i) substitute the dependent variable μ and independent variables $C_{x_1}^*, C_{x_2}^*, C_{x_3}^*, \dots, C_{x_n}^*$ into the LR model; (ii) use the LR model with the Wald-forward method to identify the independent variables $C_{y_1}^*, C_{y_2}^*, C_{y_3}^*, \dots, C_{y_k}^*$ that significantly influence the compliance probability; (iii) the significant variables $C_{y_1}^*, C_{y_2}^*, C_{y_3}^*, \dots, C_{y_k}^*$ are used as independent variables and substituted into the input layer of the RBF network model; and the μ is used as the dependent variable of the input layer to obtain a set of prediction values for the compliance probability.

2.3.2. Two-Stage Model of LR-ANN II

Model II is constructed as follows. (i) substitute the dependent variable μ and independent variables $C_{x_1}^*, C_{x_2}^*, C_{x_3}^*, \dots, C_{x_n}^*$ into the LR model and (ii) use the LR model with the Wald-forward method to identify the independent variables $C_{y_1}^*, C_{y_2}^*, C_{y_3}^*, \dots, C_{y_k}^*$ that significantly influence the compliance probability. The function of the LR model can be described as:

$$p = 1/[1 + e^{-(\beta_0 + \beta_1 C_{y_1}^* + \dots + \beta_k C_{y_k}^*)}] \quad (5)$$

which is used to obtain the prediction value of the compliance probability for each dataset; (iii) convert the compliance probability into a “negative signal” (value of 0) or “positive signal” (value 1) to produce a new dependent variable $\hat{\mu}$; (iv) the significant variables $C_{y_1}^*, C_{y_2}^*, C_{y_3}^*, \dots, C_{y_k}^*$ are used as the independent variables and substituted into the input layer of the RBF network model; and the new dependent variable $\hat{\mu}$ is used as the dependent variable of the input layer to obtain a set of prediction values for the compliance probability.

2.3.3. Two-Stage Model of LR-ANN III

Model III is constructed using the following steps: (i) substitute the dependent variable μ and independent variables $C_{x_1}^*, C_{x_2}^*, C_{x_3}^*, \dots, C_{x_n}^*$ into the LR model; (ii) use the LR model with the Wald-forward method to identify independent variables $C_{y_1}^*, C_{y_2}^*, C_{y_3}^*, \dots, C_{y_k}^*$ that significantly influence the compliance probability, where the function of the LR model is the same as in Equation (3), which is used to obtain the prediction value of the compliance probability for each dataset; (iii) the significant variables $C_{y_1}^*, C_{y_2}^*, C_{y_3}^*, \dots, C_{y_k}^*$ are used as the independent variables and substituted into the input layer of the RBF network model; and finally the prediction value of the compliance probability that is obtained in the anterior process is substituted into the RBF network model as the dependent variable of the input layer to produce a set of prediction values for the compliance probability.

2.4. Methods of Improving the Prediction Accuracy Ratio

To increase the prediction accuracy ratio of the SME credit risk for FIs in SCF, we propose four methods as follows: data normalization, collinearity diagnosis, cross validation and the optimal cutoff point.

2.4.1. Data Normalization Method

The prediction model of SME credit risk involves independent variables that have different units or degrees of variation; therefore, it is necessary to eliminate the effects of variations on the dimension and figures of the independent variables. Bekhet and Eletter [44] emphasized that data normalization can improve the network training capability such as by increasing the data handling efficiency and astringency speed. The normalization method can include a min-max algorithm, a Z-core algorithm, *etc.* In this paper, we apply the Z-core normalization algorithm which can be described as

$$C_i^* = (C_i - \bar{C})/S_i \quad (6)$$

where C_i^* are the normalized data, C_i are the source data, \bar{C} and S_i are the average value and the standard deviation of the source data, respectively.

2.4.2. Collinearity Diagnosis Method

We utilize the linear regression method to examine the phenomenon of collinearity and to exclude the variables of collinearity according to Way [45] and Goldstein [46] based on three indices: conditional index (CI), tolerance (T) and variance inflation (VIF). Because the variables with index values of $CI > 10$, $T < 0.2$ and $VIF > 10$ exhibit strong collinearity, we discard the variables whose index values reach the threshold.

2.4.3. Cross Validation Method

Zhang *et al.* [25], Lin [29], Stone [47] and Efron and Tibshirani [48] prove that the cross-validation method can be used to test and strength the predictive power of models. In this paper, we randomly divide the samples into five groups. When we test the data of one of the five groups, the data of the other four groups are used as training data for the purpose of constructing the model. We obtain the prediction accuracies of the five groups using this method. The final prediction accuracy ratio of the model is measured by the average of the five groups' test results.

2.4.4. Optimal Cutoff Point Method

To determine the cutoff point for credit risk and improve the prediction accuracy ratio of models, we adopt the optimal cutoff point approach proposed by Hosmer *et al.* [38], which is calculated based on the point of intersection of sensitivity and specificity according to Hosmer *et al.* [38] and Lin [29] (see Equation (7)). The sensitivity and specificity is calculated using Equations (8) and (9).

$$\text{optimal cutoff point} = \{[1 - \text{median}(\text{sensitivity})]^2 + [\text{median}(1 - \text{specificity})]^2\}^{1/2} \quad (7)$$

$$\text{sensitivity} = \frac{\text{number of true positives}}{\text{number of true positives} + \text{number of false negatives}} \quad (8)$$

$$\text{specificity} = \frac{\text{number of true negatives}}{\text{number of true negatives} + \text{number of false positives}} \quad (9)$$

where "sensitivity" measures the proportion of "actual positives" and is complementary to the "false negative" ratio, "specificity" measures the proportion of "negatives" and is complementary to the "false positive" ratio, and "median" is the median value.

3. Description of Data and Sampling Procedure

3.1. Assumption of Applying Supply Chain Financing (SCF)

SCF has been promoted for almost ten years in China; however, only a few SMEs, CEs and FIs cooperate to facilitate SCF in practice. Thus, we failed to gather enough empirical data in SCF from the references, interviews and surveys. Alternatively, we can use the quarterly financial and non-financial data of selected listed SMEs and CEs on a quarter-by-quarter basis because these SMEs and CEs have real trading relationships with each other. In our study, we assume that these SMEs cooperate with CEs and FIs in SCF, when SMEs are short of capital and starved for financing.

3.2. Variable Definitions

3.2.1. Dependent Variable

The dependent variable represents whether each quarterly data sample of an SME presents a high credit risk signal: a value of 0 indicates a "negative signal", which means that the SME's credit

risk is high; while a value of 1 indicates a “positive signal”, which means the SME’s credit risk is low. Following Zhu *et al.* [49], we categorize SMEs into the high and low credit risk groups, depending on whether the SME is a star special treatment (*ST) listed company. The *ST listed SME is the listed company from the Small and Medium Enterprise Board of Shenzhen Stock Exchange that is facing a delisting warning because it has suffered operating losses for two consecutive years. In other words, each quarterly data sample of *ST SMEs presents a “negative signal” in the two years before they are labeled *ST; in contrast, each quarterly data sample of non-*ST SMEs presents a “positive signal” in the past two years.

3.2.2. Independent Variables

In SCF, FIs evaluate the SME credit risks following four factors, which contain some sub-factors: applicant (SME) factor (sub-factors: capability of repayment, operational capability, profitability, development capability and credit rating), counter party (CE) factor (sub-factors: credit rating, capability of repayment and operational capability), items’ characteristic factors (sub-factors: characteristics of trade goods and characteristics of accounts receivable), and operation condition factors (sub-factors: industry status, degree of cooperation and credit worthiness of the applicant). These sub-factors are divided into 18 evaluation indices again according to the suggestions of Xiong *et al.* [33] and Zhu *et al.* [49]. These 18 evaluation indices serve as the source independent variables of the LR and ANN models. To facilitate the observation, we describe and define these independent variables as in Table 1.

Table 1. List of independent variables [49].

Indexes	Variables	Categories
C ₁	Current ratio of SME	Liquidity
C ₂	Quick ratio of SME	Liquidity
C ₃	Cash ratio of SME	Liquidity
C ₄	Working capital turnover of SME	Liquidity
C ₅	Return on equity of SME	Leverage
C ₆	Profit margin on sales of SME	Profitability
C ₇	Rate of Return on Total Assets of SME	Leverage
C ₈	Total Assets Growth Rate of SME	Activity
C ₉	Credit rating of CE	Non-financial
C ₁₀	Quick ratio of CE	Liquidity
C ₁₁	Turnover of total capital of CE	Liquidity
C ₁₂	Profit margin on sales of CE	Profitability
C ₁₃	Price rigidity, liquidation and vulnerable degree of trade goods	Non-financial
C ₁₄	Accounts receivable collection period of SME	Leverage
C ₁₅	Accounts receivable turnover ratio of SME	Leverage
C ₁₆	Industry trends of SME	Non-financial
C ₁₇	Transaction time and transaction frequency of SME	Non-financial
C ₁₈	Credit rating of SME	Non-financial

Note on Abbreviations: SME: medium-sized enterprise; CE: core enterprise.

It is noteworthy that the independent variables shown in Table 1 are financial indicators, except for C₉, C₁₃, C₁₆, C₁₇ and C₁₈. Specifically, we obtain the data samples of those 13 financial indexes from the database and collect the data of the remaining five non-financial indexes using the expert evaluation method. The non-financial indexes C₉, C₁₃, C₁₆, C₁₇ and C₁₈ are considered because they exhibit significant superiority in constructing an SCF SME credit risk evaluation index system compared with traditional SME credit risk evaluation index systems [32–35]. Therefore, as in References [32–35], the 18 independent variables fall into five categories: leverage, liquidity, profitability, activity and non-financial.

3.3. Sampling Procedure

In the empirical analysis, the samples are constituted by two sets of data. The first one is the quarterly financial and non-financial data of 77 listed SMEs, which is from the Small and Medium Enterprise Board of Shenzhen Stock Exchange from 31 March 2012 to 31 December 2013. The second one is the quarterly financial and non-financial data of 11 listed CEs which is from the Shanghai Stock Exchange and the Shenzhen Stock Exchange from the period of 31 March 2012 to 31 December 2013, respectively. In our study, 600 valid quarterly data points contained in our data set are used to test the five SME credit risk prediction models. The 11 CEs have a high degree of credit rating, while the 77 listed SMEs include 12 *ST listed companies and 65 non-*ST listed companies.

4. Experimental Results and Analysis

4.1. Experimental Results of Data Normalization

To calculate the results of the normalized data, we apply the Z-core normalization algorithm in Equation (6). In this equation, we first need to estimate the average value and standard deviation of the source data (See Table 2).

Table 2. Descriptive statistics of the mean and standard deviation of 600 samples.

Independent Variables	Observations	Mean	Std. Deviation
C ₁	600	1.794	1.665
C ₂	600	1.351	1.539
C ₃	600	0.574	0.916
C ₄	600	14.566	71.026
C ₅	600	0.049	0.074
C ₆	600	0.051	0.080
C ₇	600	0.028	0.037
C ₈	600	0.221	0.258
C ₉	600	8.155	2.702
C ₁₀	600	0.990	0.204
C ₁₁	600	0.836	0.451
C ₁₂	600	0.0419	0.027
C ₁₃	600	6.300	2.278
C ₁₄	600	76.709	49.361
C ₁₅	600	6.751	9.670
C ₁₆	600	5.695	2.012
C ₁₇	600	6.300	2.278
C ₁₈	600	5.695	2.012

4.2. Experimental Results of Collinearity Diagnosis

Because independent variables with values of $CI > 10$, $T < 0.2$ and $VIF > 10$ indicate strong collinearity, we discard variables whose values reach the threshold. The linear regression method is used for diagnosing every independent variable's threshold values of CI , T and VIF and obtaining 10 independent variables. We present the collinearity diagnosis index values of 18 independent variables and 10 reserved independent variables' new collinearity diagnosis index values in Table 3. The results of the Analysis of Variance (ANOVA) testing reveal a collinearity with significance reaching the 1% level, as shown in Table 4.

Table 3. Collinearity diagnosis index value.

Independent Variables	Original 18 Variables			Reserved 10 Variables ^d		
	T ^a	VIF ^b	CI ^c	T	VIF	CI
C ₁ [*]	0.008	126.175	1.195			
C ₂ [*]	0.006	169.747	1.230			
C ₃ [*]	0.084	11.975	1.584			
C ₄ [*]	0.911	1.098	1.792	0.933	1.072	1.173
C ₅ [*]	0.121	8.236	1.909			
C ₆ [*]	0.356	2.807	1.934	0.737	1.358	1.455
C ₇ [*]	0.108	9.236	2.088			
C ₈ [*]	0.697	1.434	2.170	0.823	1.215	1.503
C ₉ [*]	0.392	2.552	2.644	0.518	1.932	1.549
C ₁₀ [*]	0.533	1.876	3.062	0.636	1.573	1.573
C ₁₁ [*]	0.692	1.446	3.591	0.714	1.400	1.701
C ₁₂ [*]	0.445	2.200	3.827	0.522	1.915	1.985
C ₁₃ [*]	–	–	–			
C ₁₄ [*]	0.502	1.991	4.137	0.557	1.796	2.539
C ₁₅ [*]	0.437	2.289	6.995	0.458	2.184	2.767
C ₁₆ [*]	–	–	–			
C ₁₇ [*]	0.477	2.094	7.874	0.534	1.874	3.211
C ₁₈ [*]	0.471	2.124	33.233			

^a T denotes “Tolerance”; ^b VIF denotes “Variance Inflation”; ^c CI denotes “Conditional Index”; ^d There are 10 independent variables that are reserved by the collinearity diagnosis, and these reserved independent variables are used for structuring the credit risk prediction model.

Table 4. Collinearity diagnosis index value.

Model	Sum of Squares	df	Mean Square	F	Sig ^a
Regression	35.023	16	2.189	11.227	0.000
Residual Total	113.670	583	0.195		
	148.693	599			

^a Significance at 1% level.

4.3. Experimental Results of Cross Validation

We randomly divide the 600 sample points into five groups with similar sizes and distributions: groups 1, 2, 3, 4 and 5. We choose four of the groups as the training set and the remaining group as the test set. We repeat this process five times to make sure that each group has been tested.

4.4. Experimental Results of Logistic Regression (LR) Model

We use the Wald-forward method of LR for selecting significant independent variables and constructing SME credit risk prediction model. In this paper, the independent variables are excluded from the model when their significance values are larger than 0.01. The empirical LR results show that the independent variables C₆^{*}, C₉^{*}, C₁₂^{*} and C₁₄^{*} persist in the LR model (see Table 5).

Following Table 5, we first represent the LR equation as

$$\ln[p/(1-p)] = -0.282 - 0.414C_6^* + 0.866C_9^* - 0.354C_{12}^* - 0.753C_{14}^* \quad (10)$$

which indicates that the independent variables C₆^{*}, C₉^{*}, C₁₂^{*} and C₁₄^{*} have a significant influence on predicting the credit risk signals of SMEs. Furthermore, because the absolute values of the coefficients of the independent variables C₉^{*} and C₁₄^{*} are substantially larger than those of the other two independent

variables, we consider that they have a more prominent influence on predicting the credit risk signals of SMEs. The independent variable C_9^* presents a positive sign, meaning a large “Coefficient” of C_9^* , a high credit rating of CEs and a low credit risk for FIs. In contrast, the independent variable C_{14}^* presents a negative sign, meaning a large “Coefficient” of C_{14}^* carries a high credit risk for FIs.

Table 5. Significance values of independent variables.

Independent Variables	B. ^a	Sig. ^b	Situation ^c
C_4^*	0.169	0.681	Excluded
C_6^*	−0.414	0.000	Reserved
C_8^*	0.239	0.017	Excluded
C_9^*	0.866	0.000	Reserved
C_{10}^*	0.281	0.015	Excluded
C_{11}^*	−0.176	0.085	Excluded
C_{12}^*	−0.354	0.003	Reserved
C_{14}^*	−0.753	0.000	Reserved
C_{15}^*	0.123	0.762	Excluded
C_{17}^*	1.405	0.236	Excluded
Constant ^d	−0.282	0.003	Reserved

^a B. denotes “Coefficient”; ^b Sig. denotes “Significance”; ^c The independent variables are excluded when their significance values are greater than 0.01; ^d Constant denotes “intercept”.

Then, we employ the “Hosmer–Lemeshow test” for assessing the “Goodness of Fit” of the LR model. Following Hosmer *et al.* [38], in the LR model, we set the significance level to 5% and calculate the degree of freedom (DF) value using the Hosmer–Lemeshow function of SPSS (IBM Company, Chicago, USA). Based on the significance level and the DF, we calculate the critical value of the LR model as 15.507 via the “CHINV” statistics method (see Table 6). For each group testing shown in Table 6, the *p*-value is greater than 0.05, and the value of the Pearson chi-square is smaller than 15.507, suggesting that the LR model has a good fitting ability.

Table 6. Hosmer–Lemeshow test of an logistic regression model.

	Group 1	Group 2	Group 3	Group 4	Group 5
Pearson chi-square	8.808	8.810	10.830	11.199	3.068
Degree of freedom	8.000	8.000	8.000	8.000	8.000
<i>p</i> -value	0.160	0.359	0.212	0.191	0.930
Critical value	15.507	15.507	15.507	15.507	15.507

Finally, we present the optimal cutoff point for the prediction accuracy ratio of the LR model in Table 7. The experimental results show that the mean value of the “positive signal” prediction accuracy ratio is 72.8%, whereas the mean value of the “negative signal” prediction accuracy ratio is only 47.9%.

Table 7. Optimal cutoff point for the prediction accuracy ratio of the logistic regression model.

	Group 1	Group 2	Group 3	Group 4	Group 5	Mean (SD)
Optimal cutoff point ^a	0.528	0.551	0.543	0.548	0.485	0.531 (0.027)
Positive signal	52.8%	82.5%	76.2%	85.7%	66.7%	72.8% (0.133)
Negative signal	56.3%	43.9%	43.9%	42.0%	53.3%	47.9% (0.065)
Overall	54.2%	64.2%	60.8%	67.5%	60.0%	61.3% (0.050)

^a The optimal cutoff point is determined by taking the point of the intersection of the sensitivity and specificity curves.

The above model assumes that the regression coefficients are constant. Other models are also applicable. Examples include the functional-coefficient models, which assume that the coefficients are functions of a state variable [35]. This will be investigated in an ongoing project.

4.5. Experimental Results of the Artificial Neural Network (ANN) Model

In this paper, an RBF network architecture is applied to the ANN model. Therefore, we take 10 independent variables following collinearity diagnosis and use them as the input layer variables of the RBF network model. The single hidden layer of this RBF network includes 20 hidden layer neurons according to the suggestion of Wong [36]. The setting of the mean square error is 0, and the spread value is 1 in our RBF network model.

Table 8 shows that the mean value of the “positive signal” prediction accuracy ratio decreases from 72.8% of the previous model to 70.8%. However, the mean value of the “negative signal” prediction accuracy ratio is obviously increased. Moreover, the overall prediction accuracy ratio also increases from the 61.3% of the LR model to 68.8%.

Table 8. Optimal cutoff point for the prediction accuracy ratio of the artificial neural network model.

	Group 1	Group 2	Group 3	Group 4	Group 5	Mean (SD)
Optimal cutoff point ^a	0.380	0.375	0.385	0.380	0.384	0.381 (0.004)
Positive signal	61.1%	77.8%	68.3%	68.6%	78.3%	70.8% (0.073)
Negative signal	68.8%	59.6%	68.4%	80.0%	60.0%	67.4% (0.083)
Overall	64.2%	69.2%	68.3%	73.3%	69.2%	68.8% (0.032)

^a The optimal cutoff point is determined by taking the point of intersection of the sensitivity and specificity curves.

4.6. Experimental Results of Two-Stage Hybrid Model I

Table 9 shows the experimental results of the two-stage hybrid model I. The mean value of the “positive signal” prediction accuracy ratio reaches 74.9%. Moreover, the overall prediction accuracy ratio also increases from 68.8% of the ANN model to 70.2%. However, the mean of the “negative signal” prediction accuracy ratio falls behind that of the ANN model.

Table 9. Optimal cutoff point for the prediction accuracy ratio of the two-stage hybrid model I.

	Group 1	Group 2	Group 3	Group 4	Group 5	Mean (SD)
Optimal cutoff point ^a	0.459	0.410	0.437	0.452	0.384	0.428 (0.031)
Positive signal	65.3%	82.5%	73.0%	77.1%	76.7%	74.9% (0.064)
Negative signal	50.0%	68.4%	64.9%	74.0%	65.0%	64.5% (0.089)
Overall	59.2%	75.8%	69.2%	75.8%	70.8%	70.2% (0.068)

^a The optimal cutoff point is determined by taking the point of intersection of the sensitivity and specificity curves.

4.7. Experimental Results of Two-Stage Hybrid Model II

Table 10 shows the experimental results of the two-stage hybrid model II. The mean value of the “positive signal” prediction accuracy ratio and the “negative signal” prediction accuracy ratio are both dramatically enhanced. As a result, the overall prediction accuracy ratio reaches 88.5%.

Table 10. Optimal cutoff point for the prediction accuracy ratio of the two-stage hybrid model II.

	Group 1	Group 2	Group 3	Group 4	Group 5	Mean (SD)
Optimal cutoff point ^a	0.114	0.163	0.166	0.137	0.191	0.154 (0.030)
Positive signal	89.8%	92.9%	88.8%	95.5%	86.8%	90.8% (0.035)
Negative signal	93.4%	75.0%	85.0%	90.3%	75.0%	83.7% (0.085)
Overall	91.7%	87.5%	87.5%	94.2%	81.7%	88.5% (0.048)

^a The optimal cutoff point is determined by taking the point of intersection of the sensitivity and specificity curves.

4.8. Experimental Results of Two-Stage Hybrid Model III

Table 11 presents the experimental results of the two-stage hybrid model III. The results show that the mean value of the “positive signal” prediction accuracy ratio decreases from 90.8% for the two-stage hybrid model II to 86.0%. However, the mean value of the “negative signal” prediction accuracy ratio increases from 83.7% for the previous model to 88.6%. As a result, the overall prediction accuracy ratio slightly decreases from 88.5% for the two-stage hybrid model II to 87.4%. According to Bekhet and Eletter [44], Yap *et al.* [52], Kürüm *et al.* [53] and West [54], we consider that the improvement of the “negative signal” prediction accuracy ratio is more important than that of the “positive signal” prediction accuracy ratio; therefore, the two-stage hybrid model III exhibits a better credit risk prediction capability than model II.

Table 11. Optimal cutoff point for the prediction accuracy ratio of the two-stage hybrid model III.

	Group 1	Group 2	Group 3	Group 4	Group 5	Mean (SD)
Optimal cutoff point ^a	0.152	0.216	0.156	0.118	0.170	0.162 (0.036)
Positive signal	82.1%	80.0%	85.2%	92.9%	89.8%	86.0% (0.053)
Negative signal	96.9%	83.3%	83.3%	96.0%	83.6%	88.6% (0.072)
Overall	90.0%	81.7%	84.2%	94.2%	86.7%	87.4% (0.049)

^a The optimal cutoff point is determined by taking the point of intersection of the sensitivity and specificity curves.

4.9. Comparing the SME Credit Risk Prediction Accuracies of the Five Models

Hosmer *et al.* [38] argued that a better and more complete description of classification accuracy is the area under the Receiver Operating Characteristic (ROC) curve and provided general guidelines as follows:

- (1) If $ROC = 0.5$, it means no discrimination.
- (2) If $0.5 < ROC < 0.7$, it means poor discrimination.
- (3) If $0.7 < ROC < 0.8$, it means acceptable discrimination.
- (4) If $0.8 < ROC < 0.9$, it means excellent discrimination.
- (5) If $ROC \geq 0.9$, it means outstanding discrimination.

Thus, we present these five models' areas under the ROC curve in Table 12.

Table 12. Discrimination accuracies of the five models.

Models	Group 1	Group 2	Group 3	Group 4	Group 5	Mean (SD)	Discrimination Accuracy ^a
LR	0.608	0.611	0.623	0.628	0.653	0.625 (0.018)	No
ANN	0.811	0.835	0.809	0.812	0.825	0.818 (0.011)	Excellent
Hybrid I	0.751	0.796	0.764	0.764	0.819	0.779 (0.028)	Acceptable
Hybrid II	0.974	0.958	0.959	0.967	0.952	0.962 (0.009)	Outstanding
Hybrid III	0.959	0.940	0.963	0.968	0.959	0.958 (0.011)	Outstanding

^a The rule of the discrimination accuracy was proposed by Hosmer *et al.* [38].

Table 12 shows that the two-stage hybrid models II and III demonstrate outstanding performance in terms of discrimination. To compare the SME credit risk prediction accuracy ratios of the five models in detail, we illustrate the credit prediction accuracy ratios of the LR model, the ANN model and the three types of two-stage hybrid models of LR-ANN in Figure 3, where the red marks indicate the “negative signal” prediction accuracy ratios, the blue marks indicate the “positive signal” prediction accuracy ratios, the black marks indicate the overall signal prediction accuracy ratios, and the orange marks indicate the optimal cutoff point for predictive values. Panels (a–e) depict the five groups’ test results of the SME credit risk prediction accuracy ratios under the five prediction models. Panel (f) illustrates that, from the first model to the fifth model, the average prediction accuracy ratios follow a stepwise uptrend. Meanwhile, we can find that the average “negative signal” prediction accuracy ratio of the two-stage hybrid model III reaches a peak in panel (f). Specifically, the two-stage hybrid model III is better than the other four models in terms of finding the bad applications. Thus, we propose that China’s FIs should apply the two-stage hybrid model III to predict the SME credit risk in SCF.

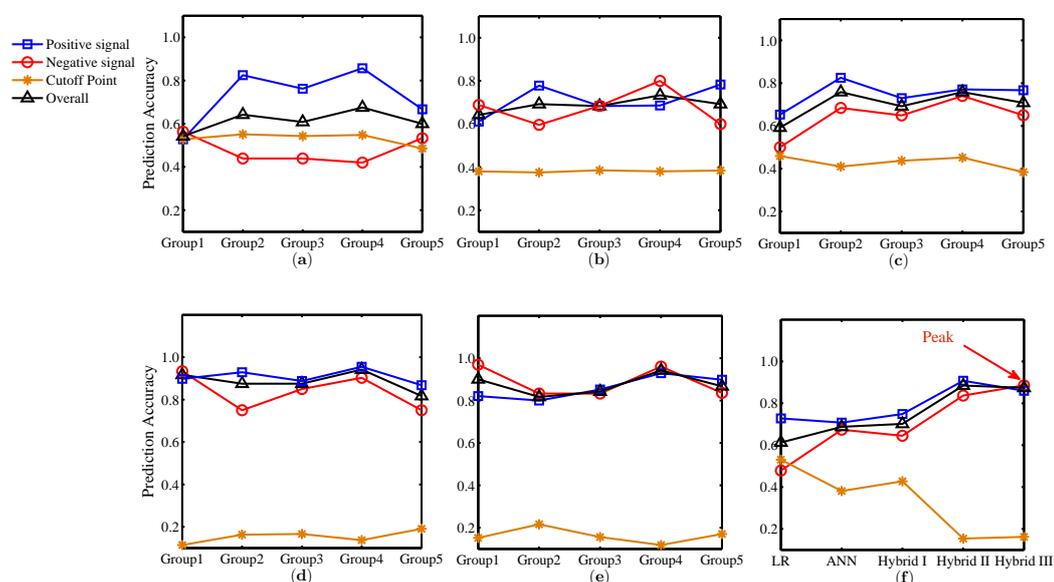


Figure 3. Comparing the SME credit risk signal prediction accuracy ratios of five models, this shows that the average “negative signal” prediction accuracy ratio of the two-stage hybrid model III reaches a peak. (a) Prediction accuracy ratios of the logistic regression; (b) prediction accuracy ratios of the artificial neural network; (c) prediction accuracy ratios of the two-stage hybrid model I; (d) prediction accuracy ratios of the two-stage hybrid model II; (e) prediction accuracy ratios of the two-stage hybrid model III; (f) mean prediction accuracy ratios of five models;

5. Conclusions

In this paper, we investigated the quarterly financial and non-financial data of 77 listed SMEs and 11 listed core enterprises CEs in China during the period of 2012–2013. Specifically, we constructed a new SME credit risk evaluation index system and five types of SME credit risk prediction models for China’s FIs in SCF. We first normalized the source data, excluded independent variables with strong collinearity, and randomly divided the samples into five groups for the purpose of model testing and construction. Then, we built five credit prediction models using the LR approach, the ANN approach and a hybrid approach.

Some basic findings for predicting SME credit risk in this study can be summarized as follows: (i) evaluating the credit risks of SMEs in SCF from four aspects, including applicants (SMEs), counter parties (CEs), items’ characteristics and operation situation; (ii) the variables C_6^* , C_9^* , C_{12}^* and C_{14}^* significantly influence the SME credit risk signals prediction accuracy ratio; and (iii) the two-stage

hybrid model III is better than the other four models in predicting “negative signals”. Because we consider that improving the ratio of bad applicant prediction accuracy is more important than improving the ratio of good applicant prediction accuracy for FIs at the present stage of China’s credit market, we affirm that the two-stage hybrid model III provides a better SME credit risk signal prediction capability than other models do.

In practice, the two-stage hybrid model III can also be used to predict other SMEs’ credit risk signals in SCF. For instance, there are eight quarterly data samples of other SMEs and CEs for two consecutive years that are not included in existing datasets; we filter out the data on the profit margin on sales of SMEs (C_6^*), the credit rating of (C_9^*), profit margin on sales of CEs (C_{12}^*) and accounts receivable collection period of SMEs (C_{14}^*) from data samples as the input layer of the two-stage hybrid model III. Then, we obtain a value of 0 or 1 from the output layer when we run the model. We define a value of 0 as a “negative signal” and a value of 1 as a “positive signal”. If we obtain eight 0 s, this indicates an extremely high credit risk, whereas if we obtain eight 1s, this indicates a relatively low credit risk; thus, the credit risk level of an SME depends on how many 0 s or 1s we obtain. In other words, the more 0 s we obtain, the more credit issues the SME has and *vice versa*. Unfortunately, only a few of China’s SMEs and CEs have cooperated on SCF over the past decade. Therefore, we have been unable to obtain adequate cases and data samples concerning SCF in practice. In future research, it will be worthwhile to find Chinese SMEs and CEs that not only have real trading relationships but also implement SCF together. This will allow China’s financial institutions to make better financing decisions in SCF.

Acknowledgments: This work was supported by the National Natural Science Foundation of China under Grant No. 71373072 and No. 71501066; the China Scholarship Council under Grant No. 201506135022; the Specialized Research Fund for the Doctoral Program of Higher Education under Grant No. 20130161110031; and the Foundation for Innovative Research Groups of the National Natural Science Foundation of China under Grant No. 71521061.

Author Contributions: All authors discussed and agreed on the ideas and scientific contributions. You Zhu and Chi Xie performed the simulations and wrote the simulation sections. You Zhu, Chi Xie, Gang-Jin Wang and Xin-Guo Yan did the mathematical modeling in the manuscript. Chi Xie, Bo Sun and Gang-Jin Wang contributed to manuscript writing and revisions.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Hofmann, E. Von der strategie bis zur finanziellen steuerung der performance in supply chains. In *Interorganizational Operations Management*; Springer Fachmedien Wiesbaden: Wiesbaden, Germany, 2013; pp. 1–20.
- More, D.; Basu, P. Challenges of supply chain finance: A detailed study and a hierarchical model based on the experiences of an Indian firm. *Bus. Process Manag. J.* **2013**, *19*, 624–647.
- Knox, A. Electronic payment: The missing link in supply chain efficiency. *J. Financ. Transform. Mark. Imperfect.* **2005**, *14*, 16–18.
- Fairchild, A. Intelligent matching: Integrating efficiencies in the financial supply chain. *Supply Chain Manag.* **2005**, *10*, 244–248.
- Hofmann, E.; Belin, O. Value proposition of SCF. In *Supply Chain Finance Solutions: Relevancel—Propositions—Market Value*; Springer: Berlin, Germany; 2011; pp. 41–45.
- Gomm, M.L. Supply chain finance: Applying finance theory to supply chain management to enhance finance in supply chains. *Int. J. Logist. Res. Appl.* **2010**, *13*, 133–142.
- Sopranzetti, B.J. Selling accounts receivable and the underinvestment problem. *Q. Rev. Econ. Financ.* **1999**, *39*, 291–301.
- Seifert, R.W.; Seifert, D. Financing the chain. *Int. Commer. Rev.* **2011**, *10*, 32–44.
- Wuttke, D.A.; Blome, C.; Henke, M. Focusing the financial flow of supply chains: An empirical investigation of financial supply chain management. *Int. J. Prod. Econ.* **2013**, *145*, 773–789.

10. Gouvêa, M.A.; Gonçalves, E.B. Credit risk analysis applying logistic regression, neural networks and genetic algorithms models. In Proceedings of the POMS 18th Annual Conference, Dallas, TX, USA, 4–7 May 2007; pp. 4–7.
11. Lahsasna, A.; Ainon, R.N.; Teh, Y.W. Credit scoring models using soft computing methods: A survey. *Int. Arab J. Inf. Technol.* **2010**, *7*, 115–123.
12. Wu, C.; Guo, Y.; Zhang, X.; Xia, H. Study of personal credit risk assessment based on support vector machine ensemble. *Int. J. Innov. Comput. Inf. Control.* **2010**, *6*, 2353–2360.
13. Burgt, M.J.V.D. Calibrating low-default portfolios, using the cumulative accuracy profile. *J. Risk Model Valid.* **2007**, *1*, 1–17.
14. Harrell, F.E.; Lee, K.L. A comparison of the discrimination of discriminant analysis and logistic regression under multivariate normality. In *Biostatistics: Statistics in Biomedical, Public Health and Environmental Sciences*; North-Holland: New York, NY, USA, 1985; pp. 333–343.
15. Altman, E.I.; Sabato, G. Effects of the New Basel Capital Accord on bank capital requirements for SMEs. *J. Financ. Serv. Res.* **2005**, *28*, 15–42.
16. Altman, E.I.; Sabato, G. Modelling credit risk for SMEs: Evidence from the US market. *Abacus J. Acc. Financ. Bus. Stud.* **2007**, *43*, 332–357.
17. Behr, P.; Güttler, A. Credit risk assessment and relationship lending: An empirical analysis of German small and medium-sized enterprises. *J. Small Bus. Manag.* **2007**, *45*, 194–213.
18. Fantazzini, D.; Figini, S. Default forecasting for small-medium enterprises: Does heterogeneity matter. *Int. J. Risk Assess. Manag.* **2009**, *11*, 38–49.
19. Fidrmuc, J.; Heinz, C. Default rates in the loan market for SMEs: Evidence from Slovakia. *Econ. Syst.* **2009**, *34*, 133–147.
20. Pederzoli, C.; Torricelli, C. A parsimonious default prediction model for Italian SMEs. *Banks Bank Syst.* **2010**, *5*, 28–32.
21. Pederzoli, C.; Thoma, G.C. Modelling credit risk for innovative firms: The role of innovation measures. *J. Financ. Serv. Res.* **2013**, *44*, 111–129.
22. Lee, T.S.; Chen, I.F. A two-stage hybrid credit scoring model using artificial neural networks and multivariate adaptive regression splines. *Expert Syst. Appl.* **2005**, *28*, 743–752.
23. Salchenberger, L.M.; Cinar, E.M.; Lash, N.A. Neural networks: A new tool for predicting thrift failures. *Decis. Sci.* **1992**, *23*, 899–916.
24. Sharda, R.; Wilson, R.L. Neural Network experiments in business-failure forecasting: Predictive performance measurement issues. *Int. J. Comput. Intell. Organ.* **1996**, *1*, 107–117.
25. Zhang, G.; Hu, M.Y.; Patuwo, B. Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *Eur. J. Oper. Res.* **1999**, *166*, 16–32.
26. Lee, T.S.; Chiu, C.C.; Lu, C.J.; Chen, I.F. Credit scoring using the hybrid neural discriminant technique. *Expert Syst. Appl.* **2002**, *23*, 245–254.
27. Chung, H.M.; Gray, P. Special section: Data mining. *J. Manag. Inf. Syst.* **1999**, *16*, 11–16.
28. Craven, M.W.; Shavlik, J.W. Using neural networks for data mining. *Future Gener. Comput. Syst.* **1997**, *13*, 211–229.
29. Lin, S.L. A new two-stage hybrid approach of credit risk in banking industry. *Expert Syst. Appl.* **2009**, *36*, 8333–8341.
30. Falavigna, G. Models for Default Risk Analysis: Focus on Artificial Neural Networks, Model Comparisons, Hybrid Frameworks. Available online: http://www.ceris.cnr.it/ceris/workingpaper/2006/WP_10_06_FALAVIGNA.pdf (accessed on 28 April 2016).
31. Deng, A.M.; Xiong, J.; Zhang, F. Order financing risk pre-warning model based on BP network. *J. Intell.* **2010**, *29*, 23–28.
32. Bai, S.Z.; Li, S. Supply chain finance risk evaluation research based on BP neural network. *Commer. Res.* **2013**, *6*, 27–31.
33. Xiong, X.; Ma, J.; Zhao, W. Credit risk analysis of supply chain finance. *Nankai Bus. Rev.* **2009**, *12*, 92–98.
34. Bai, S.B. A research into the risk early-warning of enterprise supply chain financing based on ordered logistic model. *Econ. Surv.* **2010**, *6*, 66–71.
35. Bei, Y.H.; Yang, L.; Wang, Y.H. Credit risk evaluation of car-making industry under supply chain financing mode. *Logist. Technol.* **2012**, *31*, 379–382.

36. Jiang, J.; Jiang, X.; Song, X. Weighted composite quantile regression estimation of DTARCH models. *Econ. J.* **2014**, *17*, 1–23.
37. Jiang, X.; Song, X.; Xiong, Z. Efficient and robust estimation of GARCH models. *J. Test. Eval.* **2015**, doi:10.1520/JTE20140313.
38. Hosmer, J.D.W.; Lemeshow, S.; Sturdivant, R.X. Area under the receiver operating characteristic curve. In *Applied Logistic Regression*; John Wiley & Sons: Hoboken, NJ, USA, 2013; pp. 184–197.
39. Fidell, L.S.; Tabachnick, B.G. Fundamental equations for multiple regression. In *Using Multivariate Statistics*; Harper Collins: Boston, MA, USA, 2008; pp.128–134.
40. Li, J.; Burke, E.K.; Qu R. Integrating neural networks and logistic regression to underpin hyper-heuristic search. *Knowl. Based Syst.* **2011**, *24*, 322–330.
41. Masters, T. Probabilistic neural networks I: Introduction. In *Advanced Algorithms for Neural Networks: A C++ Sourcebook*; John Wiley & Sons: New York, NY, USA, 1995; pp.112–120.
42. Gutierrez, P.A.; Segovia-Vargas, M.J.; Salcedo-Sanz, S.; Hervás-Martínez, C.; Sanchis, A.; Portilla-Figueras, J.A.; Fernández-Navarro, F. Hybridizing logistic regression with product unit and RBF networks for accurate detection and prediction of banking crises. *Omega* **2010**, *38*, 333–344.
43. Broomhead, D.S.; Lowe, D. Radial basis functions, multi-variable functional interpolation and adaptive networks. *DTIC Doc.* **1988**, *2*, 321–355.
44. Bekhet, H.A.; Eletter, S.F.K. Credit risk assessment model for Jordanian commercial banks: Neural scoring approach. *Rev. Dev. Financ.* **2014**, *4*, 20–28.
45. Way, Y. Collinearity diagnosis for a relative risk regression analysis an application to assessment of diet cancer relationship in epidemiological studies. *Stat. Med.* **1992**, *11*, 1273–1287.
46. Goldstein, R. Book reviews: Conditioning diagnostics: Collinearity and weak data in regression. *Technometrics* **1993**, *35*, 85–86.
47. Stone, M. Cross-validatory choice and assessment of statistical predictions. *J. R. Stat. Soc.* **1974**, *36*, 111–147.
48. Efron, B.; Tibshirani, R.J. Crossvalidation and other estimates of prediction. In *An Introduction to the Bootstrap*; Chapman and Hall: New York, NY, USA, 1993; pp. 237–257.
49. Zhu, Y.; Xie, C.; Wang, G.; Yan, X. Comparison of individual, ensemble and integrated ensemble machine learning methods to predict China's SMEs credit risk in supply chain finance. *Neural Comput. Appl.* **2016**, doi:10.1007/s00521-016-2304-x.
50. Jiang, J. Multivariate Functional-coefficient regression models for multivariate nonlinear times series. *Biometrika* **2014**, *101*, 689–702.
51. Wong, F.S. Time series forecasting using backpropagation neural networks. *Neurocomputing* **1991**, *2*, 147–159.
52. Yap, B.W.; Ong, S.H.; Husain, N.H.M. Using data mining to improve assessment of credit worthiness via credit scoring models. *Expert Syst. Appl.* **2011**, *38*, 13274–13283.
53. Kürüm, E.; Yildirak, K.; Weber, G.W. A classification problem of credit risk rating investigated and solved by optimization of ROC curve. *Cent. Eur. J. Oper. Res.* **2012**, *20*, 529–557.
54. West, D. Neural network credit scoring models. *Comput. Oper. Res.* **2000**, *27*, 1131–1152.



© 2016 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC-BY) license (<http://creativecommons.org/licenses/by/4.0/>).