Factorial Network Models To Improve P2P Credit Risk Management

Ahelegbey, Daniel Felix and Giudici, Paolo and Hadji-Misheva, Branka

Boston University - Department of Mathematics and Statistics, University of Pavia - Faculty of Economics, ZHAW University of Applied Sciences

26 February 2019

Factorial Network Models To Improve P2P Credit Risk Management

Daniel Felix Ahelegbey[a,*], Paolo Giudici[1], Branka Hadji-Misheva[c]

[a] *Department of Mathematics and Statistics, Boston University, USA*
[b] *Department of Economics and Management, University of Pavia, Italy*
[c] *ZHAW University of Applied Sciences, Zurich, Switzerland*

Abstract

This paper investigates how to improve statistical-based credit scoring of SMEs involved in P2P lending. The methodology discussed in the paper is a factor network-based segmentation for credit score modeling. The approach first constructs a network of SMEs where links emerge from comovement of latent factors, which allows us to segment the heterogeneous population into clusters. We then build a credit score model for each cluster via lasso logistic regression. We compare our approach with the conventional logistic model by analyzing the credit score of over 15000 SMEs engaged in P2P lending services across Europe. The result reveals that credit risk modeling using our network-based segmentation achieves higher predictive performance than the conventional model.

Keywords: Credit Risk, Factor models, Fintech, Peer-to-Peer lending, Credit Scoring, Lasso, Segmentation

1. Introduction

Issuance of loans by traditional financial institutions, such as banks, to other firms and individuals, is often associated with major risks. The failure of loan recipients to honor their obligation at the time of maturity leaves the banks vulnerable and affects their operations. The risk associated with such transactions is referred to as credit risk. It is well known that some percentage of these non-performing loans are eventually imputed to economic losses. To minimize such risk exposures, various methods have been extensively discussed in the credit risk literature to enable credit-issuing institutions to undertake a thorough assessment to classify loan applicants into risky and non-risky customers. Some of these methods range from logistic and linear probability models to decision trees, neural networks and support vector machines. A conventional individual-level reduced-form approach is the credit scoring model which attributes a score of credit-worthiness to each loan applicant based on the available history of their financial characteristics. See Altman (1968) for some pioneer works on corporate bankruptcy prediction models using accounting-based measures as variables. For a comprehensive review on credit scoring models, see Alam et al. (2010).

Recent advancements gradually transforming the traditional economic and financial system is the emergence of digital-based systems. Such systems present a paradigm shift from

*Corresponding author at: Department of Mathematics and Statistics, Boston University, USA.
Email addresses: dfkahey@bu.edu (Daniel Felix Ahelegbey), paolo.giudici@unipv.it (Paolo Giudici), branka.hadjimisheva01@universitadipavia.it (Branka Hadji-Misheva)

traditional infrastructural systems to technological (digital) systems. Financial technological ("FinTech") companies are gradually gaining grounds in major developed economies across the world. The emergence of Peer-to-Peer (P2P) platforms is a typical example of a FinTech system. The P2P platform aims at facilitating credit services by connecting individual lenders with individual borrowers without the interference of traditional banks as intermediaries. Such platform serves as a digital financial market and an alternative to the traditional physical financial market. P2P platforms significantly improve the customer experience and the speed of the service and reduce costs to both individual borrowers and lenders as well as small business owners. Despite the various advantages, P2P systems inherit some of the challenges of traditional credit risk management. In addition, they are characterized by asymmetry of information and by a strong interconnectedness among their users (see e.g. Giudici and Hadji-Misheva, 2017) that makes distinguishing healthy and risky credit applicants difficult, thus affecting credit issuers. There is, therefore, a need to explore methods that can help improve credit scoring of individual or companies that engage in P2P credit services.

This paper investigates how to improve statistical-based credit scoring small and medium enterprises (SMEs) involved in P2P lending. The methodology discussed in the paper is a factor network-based segmentation for credit score modeling. The approach first constructs a network of SMEs where links emerge from the comovement of the latent factors that drive the observed data on individual/firm financial characteristics. The network structure then allows us to segment the heterogeneous population into two sub-groups of connected and non-connected clusters. We then build a credit score model for each sub-population via lasso logistic regression.

The contribution to the literature of this paper is manifold. Firstly, we extend the ideas contained in the factor network-based classification of Ahelegbey et al. (2019) to a more realistic setting, characterized by a large number of observations which, when links between them are the main object of analysis, becomes extremely challenging.

Secondly, we extend the network-based scoring model proposed in Giudici and Hadji-Misheva (2017) to a setting characterized by a large number of explanatory variables. The variables are selected via lasso regularization (Tibshirani, 1996; Trevor et al., 2009) and, then, summarized by factor scores. Thus, we contribute to network-based models for credit risk quantification. Network models have been shown to be effective in gauging the vulnerabilities among financial institutions for risk transmission (see Ahelegbey et al., 2016a; Battiston et al., 2012; Billio et al., 2012; Diebold and Yilmaz, 2014), and a scheme to complement micro-prudential supervision with macro-prudential surveillance to ensure financial stability (see IMF, 2011; Moghadam and Viñals, 2010; Viñals et al., 2012). Recent application of networks have been shown to improve loan default predictions and capturing information that reflects underlying common features (see Ahelegbey et al., 2019; Letizia and Lillo, 2018).

Thirdly, our empirical application contributes to modeling credit risk in SMEs particularly engaged in P2P lending. For related works on P2P lending via logistic regression, see Andreeva et al. (2007); Barrios et al. (2014); Emekter et al. (2015); Serrano-Cinca and Gutiérrez-Nieto (2016). We model the credit score of over 15000 SMEs engaged in P2P credit services across Southern Europe. We compare the performance of our network-based segmentation credit score model (NetSeg-CSM) with the conventional single credit score model (Single-CSM). We show via our empirical results that our network-based segmentation presents a more efficient scheme that achieves higher performance than the conventional approach.

The organization of the paper is as follows. Section 2 presents the methodology for factor network-based segmentation and the model for credit scoring. Section 3 discusses

the application to SME and the comparison of our methodology against the conventional approach.

2. Methodology

We present the formulation and inference of a latent factor network to improve credit scoring and model estimation. Our objective is to analyze the characteristics of the borrowers to build a model that predicts the likelihood of their default.

2.1. Logistic Model

Let Y be a vector of independent observations of the loan status of n firms, such that $Y_i = 1$ if firm-i has defaulted on its loan obligation, and zero otherwise. Furthermore, let $X = \{X_{ij}\}$, $i = 1, \ldots, n$, $j = 1, \ldots, p$, be a matrix of n observations with p financial characteristic variables or predictors. The conventional parameterization of the conditional distribution of Y given X is the logistic model with log-odds ratio given by

$$\log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + X_i\beta \tag{1}$$

where $\pi_i = P(Y_i = 1|X_i)$, β_0 is a constant term, $\beta = (\beta_1, \ldots, \beta_p)'$ is a $p \times 1$ vector of coefficients and X_i is the i-th row of X.

2.2. Decomposition of Data Matrix by Factors

The dataset X can be considered as points of n-institutions in a p-dimensional space. It can also be interpreted at observed outcomes driven by some underlying firm characteristics. More specifically, X can be expressed as a factor model given by

$$X \;=\; FW' + \varepsilon \tag{2}$$

where F is $n \times k$ matrix of latent factors, W is $p \times k$ matrix of factor loadings, ε is $n \times p$ matrix of errors, and ε and F are mutually independent matrix normal random variables. In the context of our application, we set k to be the number of factors that account for a large percentage (approximately 95%) of the variation in X.

2.3. Factor Network-Based Segmentation

We present the construction of network structure for the segmentation of the population. Following the literature on graphical models (see Ahelegbey et al., 2016a,b; Carvalho and West, 2007; Eichler, 2007), we represent the network structure as an undirected binary matrix, $G \in \{0,1\}^{n\times n}$, where G_{ij} represents the presence or absence of a link between nodes i and j. We construct G via similarity of the latent firm characteristics, such that $G_{ij} = 1$ if the latent coordinates of firm-i are strongly related to firm-j, and zero otherwise.

Given the latent factors matrix, F, we construct a network where the marginal probability of a link between nodes-i and j by

$$\gamma_{ij} \;=\; P(G_{ij} = 1|F) \;=\; \Phi[\theta + (FF')_{ij}] \tag{3}$$

where $\gamma_{ij} \in (0,1)$, Φ is the standard normal cumulative density function, $\theta \in \mathbb{R}$ is a network density parameter, and $(FF')_{ij}$ is the i-th row and the j-th column of FF'. Under the

assumption that G is undirected, it follows that $\gamma_{ij} = P(G_{ij} = 1|F) = P(G_{ji} = 1|F) = \gamma_{ji}$. We validate the link between nodes-i and j in G by

$$G_{ij} \;=\; \mathbf{1}(\gamma_{ij} > \gamma) \tag{4}$$

where $\mathbf{1}(\gamma_{ij} > \gamma)$ is the indicator function, i.e., unity if $\gamma_{ij} > \gamma$ and zero otherwise, and $\gamma \in (0, 1)$ is a threshold parameter. By definition, the parameters θ and γ control the density of G. Following Ahelegbey et al. (2019), we set $\theta = \Phi^{-1}(\frac{2}{n-1})$. To broaden the robustness of the results, we compare different threshold values of $\gamma = \{0.01, 0.05, 0.1\}$ to capture a sparse but closely connected community.

2.4. Estimating High-Dimensional Logistic Models

When estimating high-dimensional logistic models with relatively large number of predictors, there is the tendency to have redundant explanatory variables. Thus, to construct a predictable model, there is the need to select the subset of predictors that explains a large variation in the probability of defaults. Several variable selection methods have been discussed and applied for various regression models. In this paper, we consider the Lasso approach (Tibshirani, 1996) for logistic regressions (Trevor et al., 2009). The objective of the Lasso logistic regression is to solve a penalized log-likelihood function given by

$$\mathcal{L}_\lambda = \sum_{i=1}^{n} \Big[Y_i(\beta_0 + X_i\beta) - \log\left(1 + \exp(\beta_0 + X_i\beta)\right) \Big] - \lambda \sum_{j=0}^{p} |\beta_j| \tag{5}$$

where n is the number of observations, p the number of predictors, and λ is the penalty term, such that large values of λ shrinks a large number of the coefficients towards zero.

3. Application

3.1. Data Description

To illustrate the effectiveness of the application of factor network methodology in credit scoring analysis, we obtained data from the European External Credit Assessment Institution (ECAI) on 15045 small-medium enterprises engaged in Peer-to-Peer lending on digital platforms across Southern Europe. The observation on each institution is composed of 24 financial characteristic ratios constructed from official financial information recorded in 2015. Table 1 presents a summary of the financial ratios in the dataset in terms of the mean of the institutions grouped according to active and defaulted SME's. In all, the data consists of 1,632 (10.85%) defaulted institutions and 13,413 (89.15%) non-defaulted companies.

3.2. Decomposition of the Observed Data Matrix by Factors

To decompose observed data matrix, X, to obtain the underlying factors that drive the observed financial characteristics, we perform a singular value decomposition given by, $X = UDW'$, where D is a diagonal matrix of non-negative and decreasing singular values and U and W are orthonormal. U is $n \times p$, D is $p \times p$ and W is $p \times p$. We obtain the underlying factor matrix by, $F = UD$, where F is a projection of X unto the eigenspace spanned by U. We retain the first $k < p$ eigenvalues that are associated with the largest variance matrix. Table 2 shows the eigenvalues of the singular value decomposition to determine the factors to retain. From the table, we retain the first 17 eigenvalues since they explain about 95% of the total variance in X.

Var	Formula (Description)	Active Mean	Defaulted Mean
V_1	(Total Assets - Shareholders Funds)/Shareholders Funds	8.87	9.08
V_2	(Longterm debt + Loans)/Shareholders Funds	1.25	1.32
V_3	Total Assets/Total Liabilities	1.51	1.07
V_4	Current Assets/Current Liabilities	1.6	1.06
V_5	(Current Assets - Current assets: stocks)/Current Liabilities	1.24	0.79
V_6	(Shareholders Funds + Non current liabilities)/Fixed Assets	8.07	5.99
V_7	EBIT/Interest paid	26.39	-2.75
V_8	(Profit (loss) before tax + Interest paid)/Total Assets	0.05	-0.13
V_9	P/L after tax/Shareholders Funds	0.02	-0.73
V_{10}	Operating Revenues/Total Assets	1.38	1.27
V_{11}	Sales/Total Assets	1.34	1.25
V_{12}	Interest Paid/(Profit before taxes + Interest Paid)	0.21	0.08
V_{13}	EBITDA/Interest Paid	40.91	5.71
V_{14}	EBITDA/Operating Revenues	0.08	-0.12
V_{15}	EBITDA/Sales	0.09	-0.12
V_{16}	Constraint EBIT	0.13	0.56
V_{17}	Constraint PL before tax	0.16	0.61
V_{18}	Constraint Financial PL	0.93	0.98
V_{19}	Constraint P/L for period	0.19	0.64
V_{20}	Trade Payables/Operating Revenues	100.3	139.30
V_{21}	Trade Receivables/Operating Revenues	67.59	147.12
V_{22}	Inventories/Operating Revenues	90.99	134.93
V_{23}	Total Revenue	3557	2083
V_{24}	Industry Classification on NACE code	4566	4624
	Total number of institutions (%)	13413 (89.15%)	1632 (10.85%)

Table 1: List and Summary Statistics of the variables in our sample.

3.3. Factor Network Analysis

For purposes of graphical representations and to keep the companies name anonymous, we report the estimated network by representing the group of institutions with color-codes. The defaulted companies are represented in red color code, and non-defaulted companies in green color code (see Figure 1). Table 3 reports the summary statistics of the estimated network in terms of the default-status composition of the SMEs. For robustness purposes, we compare the results obtained with a default threshold value $\gamma = 0.05$ against $\gamma = 0.10$ and $\gamma = 0.01$.

The result for the default threshold value $\gamma = 0.05$ of Table 3 shows that the connected sub-population is is composed of 4305 companies which constitute 28.6% of the full sample. The non-connected sub-population, on the other hand, is composed of 10740 (71.4%). The percentage of the defaulted class of companies are 22.4% and 6.2% among the connected- and non-connected sub-population, respectively. We notice that higher threshold values (say $\gamma = 0.1$) decrease (increase) the total number of connected (non-connected) sub-population and vice versa. Such higher threshold values also lead to a lower (higher) number of defaulted class of connected (non-connected) SMEs but (and) constituting a higher percentage of the defaulted population. The reverse is also true for $\gamma = 0.01$.

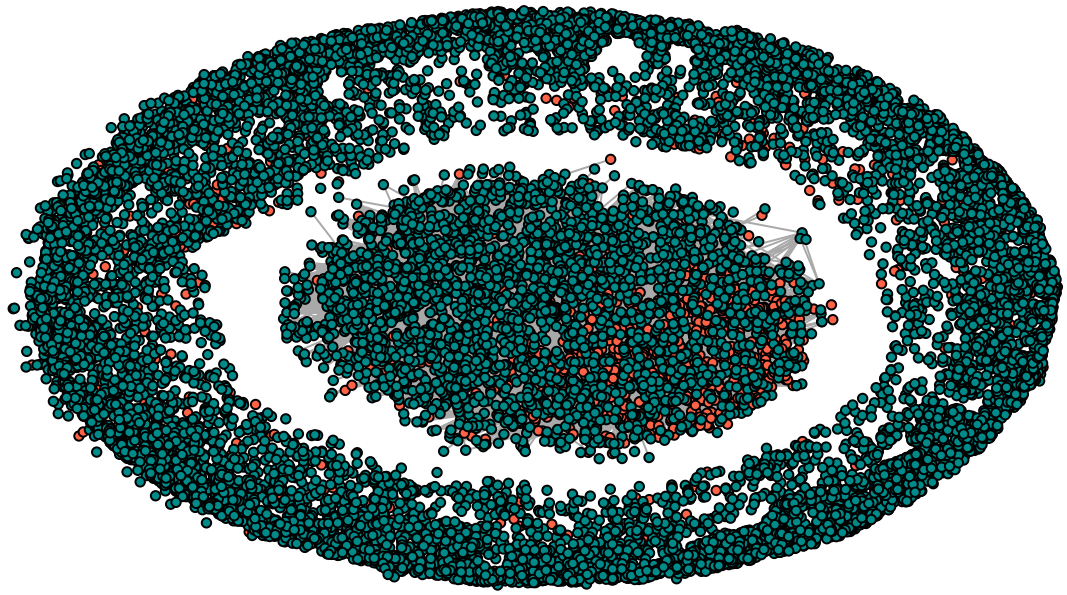
Figure 1 presents the graphical representation of the estimated factor network with the sub-population of defaulted and non-defaulted companies color coded as red and green, respectively. Figure 1a shows the structural representation of both connected and non-connected sub-population while Figure 1b depicts the structure of connected sub-population only.

No.	Eigenvalue	Variance Explained (%)	Cumulative (%)
1	5.18	21.60	21.60
2	2.58	10.73	32.33
3	2.50	10.41	42.74
4	1.60	6.69	49.42
5	1.42	5.92	55.34
6	1.30	5.40	60.74
7	1.16	4.82	65.55
8	1.09	4.56	70.11
9	0.99	4.11	74.22
10	0.93	3.88	78.10
11	0.80	3.35	81.45
12	0.79	3.31	84.76
13	0.75	3.11	87.87
14	0.56	2.35	90.22
15	0.53	2.21	92.43
16	0.51	2.12	94.55
17	0.43	1.80	96.35
18	0.37	1.54	97.89
19	0.17	0.69	98.58
20	0.11	0.47	99.05
21	0.09	0.36	99.41
22	0.07	0.27	99.68
23	0.06	0.26	99.94
24	0.01	0.06	100.00

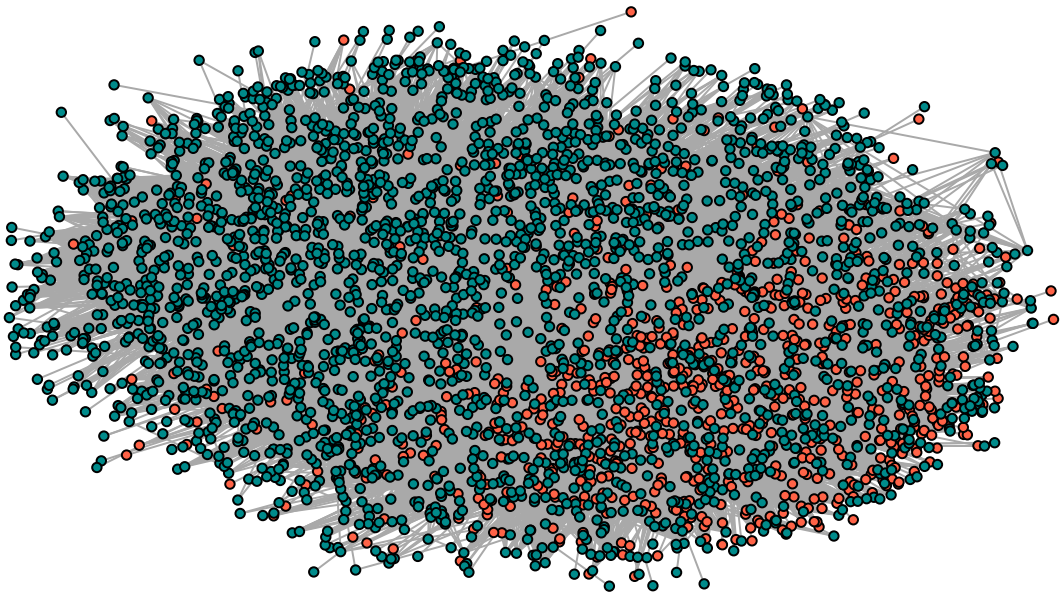
Table 2: The eigenvalues of the singular value decomposition to determine the factors to retain.

Threshold	Status	Conn-Sub	Non-Conn-Sub
$\gamma = 0.01$	Default	1,214 - 17.8%	418 - 5.1%
	Non-Default	5,602 - 82.2%	7,811 - 94.9%
	Total	6,816 - 45.3%	8,229 - 54.7%
$\gamma = 0.05$	Default	964 - 22.4%	668 - 6.2%
	Non-Default	3,341 - 77.6%	10,072 - 93.8%
	Total	4,305 - 28.6%	10,740 - 71.4%
$\gamma = 0.1$	Default	816 - 24%	816 - 7%
	Non-Default	2,580 - 76%	10,833 - 93%
	Total	3,396 - 22.6%	11,649 - 77.6%

Table 3: Summary statistic of connected and non-connected sub-population obtained from the factor network-based segmentation for threshold values of $\gamma = \{0.01, 0.05, 0.1\}$.



(a) Network Structure of All Institutions



(b) Network of Connected Component

Figure 1: A graphical representation of the estimated factor network. (1a) shows the structural representation of the factor network for default threshold $\gamma = 0.05$, and (1b) depicts the connected sub-population only. The nodes represent the companies with the red-color coded nodes representing a defaulted class of companies while green-color coded nodes denote the non-defaulted class of companies.

3.4. Lasso Logistic Credit Score Modeling

To model the credit score of the companies represents in the dataset, we employ Lasso logistic regression. To achieve this, we apply ten-fold cross-validation to select the regularization parameter ($\lambda.min$) that minimizes the cross-validated mean squared error. Our preliminary analysis shows that selecting the best model based on the "one-standard-error" rule (i.e., $\lambda.1se$) produces a model that is too restrictive in the sense that it sometimes renders

all the regressors insignificant. Thus, in this application, we rather choose $\lambda.min$ over $\lambda.1se$.

Table 4 presents a comparison of the selected variables for the coventional model (Single-CSM), the connected sub-population model (Net-Seg-CSM (C)), and the non-connected sub-population model (Net-Seg-CSM (NC)). The table shows the result of Net-Seg-CSM (C) and Net-Seg-CSM (NC) for the threshold value $\gamma = 0.1$. We observed a significant difference in the number of selected explanatory variables for the Single-CSM and the Net-Seg-CSM models. More precisely, the Single-CSM models the credit score of a given company by using almost all the information on the financial characteristic variables captured in the dataset. The Net-Seg-CSM, on the other hand, uses a significantly lower number of financial characteristic variables to model the credit scores. Thus, the factor network-based segmentation credit score framework is more parsimonious than the conventional full population credit score model, and this helps interpretability.

	Single-CSM	Net-Seg-CSM (C)	Net-Seg-CSM (NC)
(Intercept)	-1.961	-1.811	-1.126
V_1	0.003	0	0.002
V_2	0	0.009	0
V_3	-0.562	-0.348	-1.237
V_4	-0.298	-0.106	-0.437
V_5	0.003	0	0
V_6	0.003	0	0.005
V_7	0.004	0	0
V_8	-2.683	-2.322	0.519
V_9	-0.045	0.042	-0.576
V_{10}	-0.145	0.023	0
V_{11}	0.202	0	0.035
V_{12}	0.060	0.038	0.033
V_{13}	-0.003	0	0
V_{14}	-0.177	-0.400	0
V_{15}	-0.360	-0.174	0
V_{16}	0.155	0.726	0
V_{17}	0.538	0.412	0.398
V_{18}	0.167	0.256	0.025
V_{19}	0.594	0.065	0.492
V_{20}	0.0001	0	0.001
V_{21}	0.002	0.001	0.003
V_{22}	0.001	0	0.001
V_{23}	-0.00003	-0.00001	-0.00004
V_{24}	-0.00000	-0.00003	0.00001

Table 4: Comparing the Lasso logistic model estimated coefficients of the competing models. Single-CSM is the single credit score model, Net-Seg-CSM(C) is the network-based segmentation credit score model for connected sub-population, and Net-Seg-CSM(NC) is the network-based segmentation credit score models for non-connected sub-population for the threshold value (Net-Seg-CSM) for threshold value $\gamma = 0.1$.

3.5. Comparing Default Predicting Accuracy

We now compare the default prediction accuracy of the models in terms of the standard area under the curve (AUC) derived from the receiver operator characteristic (ROC) curve. The AUC depicts the true positive rate (TPR) against the false positive rate (FPR) depending on some threshold. TPR is the number of correct positive predictions divided by the total number of positives. FPR is the ratio of false positives predictions overall negatives. See Figure 2 for the plot of the ROC curve for the competing methods.

We analyzed the performance of the models by splitting the sample into 70% training and 30% testing sample. We report in Table 5 the confusion matrices obtained from the prediction of the probability of default via the conventional single credit score model (Single-CSM) and the network-based segmented credit score models (Net-Seg-CSM) for threshold values of $\gamma = \{0.01, 0.05, 0.1\}$. The confusion matrix that presents the number of true positives, false positives, true negative and false negative outcomes under the competing methods.

	Single-CSM		Net-Seg-CSM ($\gamma = 0.01$)		Net-Seg-CSM ($\gamma = 0.05$)		Net-Seg-CSM ($\gamma = 0.1$)	
	True		True		True		True	
Pred	0	1	0	1	0	1	0	1
0	3,996	347	3,970	367	3,960	361	3972	382
1	49	122	60	117	56	138	37	123

Table 5: Confusion matrices obtained from the prediction of the probability of default from the conventional single credit score model (Single-CSM) and the network-based segmented credit score models (Net-Seg-CSM) for threshold values of $\gamma = \{0.01, 0.05, 0.1\}$.

	Single-CSM	Net-Seg-CSM ($\gamma = 0.01$)	Net-Seg-CSM ($\gamma = 0.05$)	Net-Seg-CSM ($\gamma = 0.1$)
AUC	0.8089	0.8331	0.8220	0.8323

Table 6: Comparing model performance of the prediction of the probability of default from the conventional single credit score model (Single-CSM) and the network-based segmented credit score models (Net-Seg-CSM) for threshold values of $\gamma = \{0.01, 0.05, 0.1\}$.

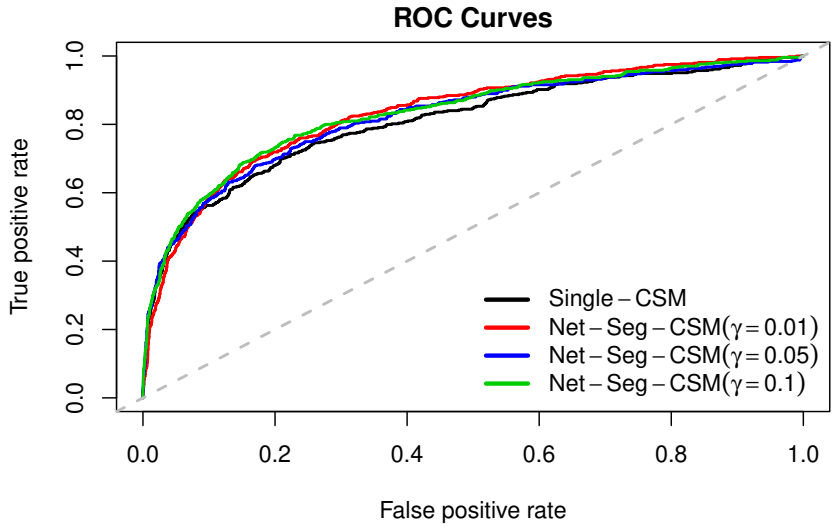


Figure 2: Plot of the ROC curves of the conventional single credit score model (Single-CSM) and the network-based segmented credit score models (Net-Seg-CSM) for threshold values of $\gamma = \{0.01, 0.05, 0.1\}$.

The comparison of the ROC curves from the competing methods shows that the Single-CSM (in black) lies below the rest. Clearly, the curves of Net-Seg-CSM ($\gamma = 0.01$) and Net-Seg-CSM ($\gamma = 0.1$) depicted in red and green, respectively, seems to dominate the others. However, none of these two completely dominate the other, depending on the cut-off that is

chosen to predict a company to be default or active. The summary of the area under the ROC curve reported in Table 6 shows that Net-Seg-CSM ($\gamma = 0.01$) is ranked first, followed by Net-Seg-CSM ($\gamma = 0.1$). The lowest AUC is obtained by the Single-CSM. Overall, in terms of default predictive accuracy, the result of the AUC shows the Net-Seg-CSM outperforms the Single-CSM, on average by two percentage points. An advantage that can be further increase considering as cut-off the observed default percentages, different in the two samples.

In conclusion, our proposed factor network approach to credit score modeling presents an efficient framework to analyze the interconnections among the borrowers of a peer to peer platform and provides a way to segment a heterogeneous population into clusters with homogeneous characteristics. The results show that the lasso logistic model for credit scoring leads to a better identification of the significant set of relevant financial characteristic variables, thereby, producing a more interpretable model, especially when combined with the segmentation of the population via the factor network-based approach. We also find evidence of an improvement in the default predictive performance of our model compared to the conventional approach.

4. Conclusion

This paper improves credit risk management of SMEs engaged in P2P credit services by proposing a factor network-based approach to segment a heterogeneous population into a cluster of homogeneous sub-populations and estimating a credit score model on the clusters using a Lasso logistic model.

We demonstrate the effectiveness of our approach through empirical applications to analyze the probability of default of over 15000 SMEs involved in P2P lending across Europe. We compare the results from our model with the one obtained with standard single credit score methods.

We find evidence that our factor network approach helps is obtain sub-population clusters such that the resulting models associated with these clusters are more parsimonious than the conventional full population approach, leading to better interpretability and to an improved default predictive performance.

References

Acemoglu D, Ozdaglar A, Tahbaz-Salehi A. 2015. Systemic Risk and Stability in Financial Networks. *American Economic Review* **105**: 564–608.

Ahelegbey D, Giudici P, Hadji-Misheva B. 2019. Latent Factor Models For Credit Scoring in P2P Systems. *Physica A: Statistical Mechanics and its Applications* **522**: 112–121.

Ahelegbey DF, Billio M, Casarin R. 2016a. Bayesian Graphical Models for Structural Vector Autoregressive Processes. *Journal of Applied Econometrics* **31**: 357–386.

Ahelegbey DF, Billio M, Casarin R. 2016b. Sparse Graphical Vector Autoregression: A Bayesian Approach. *Annals of Economics and Statistics* **123/124**: 333–361.

Alam M, Hao C, Carling K. 2010. Review of the literature on credit risk modeling: development of the past 10 years. *Banks and Bank Systems* **5**: 43–60.

Altman EI. 1968. Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankrupcy. *The Journal of Finance* **23**: 589–609.

Andreeva G, Ansell J, Crook J. 2007. Modelling Profitability Using Survival Combination Scores. *European Journal of Operational Research* **183**: 1537–1549.

Barrios LJS, Andreeva G, Ansell J. 2014. Monetary and Relative Scorecards to Assess Profits in Consumer Revolving Credit. *Journal of the Operational Research Society* **65**: 443–453.

Battiston S, Gatti DD, Gallegati M, Greenwald B, Stiglitz JE. 2012. Liaisons Dangereuses: Increasing Connectivity, Risk Sharing, and Systemic Risk. *Journal of Economic Dynamics and Control* **36**: 1121–1141.

Billio M, Getmansky M, Lo AW, Pelizzon L. 2012. Econometric Measures of Connectedness and Systemic Risk in the Finance and Insurance Sectors. *Journal of Financial Economics* **104**: 535 – 559.

Carvalho CM, West M. 2007. Dynamic Matrix-Variate Graphical Models. *Bayesian Analysis* **2**: 69–98.

Diebold F, Yilmaz K. 2014. On the Network Topology of Variance Decompositions: Measuring the Connectedness of Financial Firms. *Journal of Econometrics* **182**: 119–134.

Eichler M. 2007. Granger Causality and Path Diagrams for Multivariate Time Series. *Journal of Econometrics* **137**: 334–353.

Emekter R, Tu Y, Jirasakuldech B, Lu M. 2015. Evaluating Credit Risk and Loan Performance in Online Peer-to-Peer (P2P) Lending. *Applied Economics* **47**: 54–70.

Giudici P, Hadji-Misheva B. 2017. Scoring Models for P2P Lending Platforms: A Network Approach. Working paper, University of Pavia.

IMF. 2011. Global Financial Stability Report: Grappling with Crisis Legacies. Technical report, World Economic and Financial Services.

Letizia E, Lillo F. 2018. Corporate Payments Networks and Credit Risk Rating. Working paper, https://arxiv.org/abs/1711.07677.

Moghadam R, Viñals J. 2010. Understanding Financial Interconnectedness. Mimeo, International Monetary Fund.

Serrano-Cinca C, Gutiérrez-Nieto B. 2016. The Use of Profit Scoring as an Alternative to Credit Scoring Systems in Peer-to-Peer Lending. *Decision Support Systems* **89**: 113–122.

Tibshirani R. 1996. Regression Shrinkage and Selection via the LASSO. *Journal of the Royal Statistical Society. Series B* **58**: 267–288.

Trevor H, Robert T, JH F. 2009. The Elements of Statistical Learning: Data Mining, Inference and Prediction.

Viñals J, Tiwari S, Blanchard O. 2012. *The IMF'S Financial Surveillance Strategy*. International Monetary Fund.