ICIC Express Letters
Part B: Applications
Volume **7**, Number **7**, July **2016**

 ISSN 2185-2766
pp. **1491–1496**

COMPARISON BETWEEN BP NEURAL NETWORK AND STRUCTURAL EQUATION MODEL IN C2C E-COMMERCE CREDIT EVALUATION

Dan Bai and Min Zuo*

College of Economics and Management
Dalian University
No. 10, Xuefu Ave., Dalian 116622, P. R. China
baidan@dlu.edu.cn; *Corresponding author: 731053091@qq.com

Received December 2015; accepted March 2016

Abstract. *During generating simulated data sets and case study, this paper uses the consistency index C to compare the performance of back propagation (BP) neural network and structural equation model (SEM) in the consumer to consumer (C2C) E-commerce credit evaluation. The study has shown that for the small sample size which has low missing ratio and data which has few restrictions, BP neural network has better advantage than structural equation model, which is worth further popularizing in C2C E-commerce credit evaluation.*
Keywords: BP neural network, Structural equation model, C2C E-commerce credit evaluation, Comparative analysis

1. **Introduction.** By influencing Chinese economic and social development and changing people's lifestyle, the Internet became a key industry. However, the trust question of Internet has not been solved completely which limited the development of C2C E-commerce. And how to improve the Internet users' trust in C2C E-commerce is still an important question.

2. **Methods Introduction and Problem Statement.** Back propagation (BP) neural network is a back propagation algorithm multilayer feed-forward network, and it is one of the most widely used neural network models. BP network can learn and store a large amount of input-output mode mapping without prior describing this mathematical equation. Its learning rule is the steepest descent method, constantly adjusting the network's weights and thresholds by reverse spread until the squared error is the smallest.

Structural equation model (SEM) as a multivariate statistical method, which is based on the factor analysis, path analysis and regression analysis, can estimate a set of observed variables and latent variables as well as analyzing the relationship between the variables [1].

Most scholars' studies are more concentrated in one aspect, and the comparative analysis between BP neural network and SEM has deficiency. Liu and Xiang use SEM to study the E-commerce satisfaction [2]. Jiang and Hu use the method of SEM to build C2C E-commerce trust model and analyse the affecting factors [3]. He et al. analyse the individual electronic commerce credit problem based on BP neural network [4]. Only Zhao and Wan made a comparative analysis of the two methods in the case study of business performance and simply explored the complementarity between the two methods based on the analysis results [5].

However, in the field of C2C E-commerce credit evaluation, there is no the comparative analysis between BP neural network and SEM. Which method is good for C2C

E-commerce credit evaluation and which method has a better performance in evaluating the credit of the C2C E-commerce? These questions are necessary to be solved. So the authors combine theoretical and practical application to compare the performance of BP neural network and SEM. It is the first time to contrast the performance of the two methods in the field of C2C E-commerce credit evaluation.

3. Simulation and Case Study.

3.1. Simulation design.

Simulation data include: different sample sizes, different percentages of missing and different latent variable assumptions, different data flows.

Simulation missing data: the simulated study generates the right missing data, and missing data distribution is a uniform distribution $U(0, u)$. Firstly it generates n (n is the sample size) data T_i, submitting the distribution of $U(0, u)$ and the missing proportion is by iterative simulation, then it counts the value of the u. And secondly the SAS program generates n missing data L_i. Finally, according to the definition of the right missing, $t_i = \min(T_i, L_i)$, it gives the n observations. If $T_i \leq L_i$, $\delta_i = 1$, otherwise, generating samples which contain the missing data (t_i, δ_i) [6].

Simulations date collection: it generates the random data using RANNOR (seed) function which obeys normal distribution $N(0, 1)$, and generates the random variable using RANNOR (seed, n, p) function which obeys Bernoulli Lee distributed $B(n, p)$ [7].

Using the above two functions to generate random simulation data, we ensure that the data are random and objective.

Simulation evaluation: because the predictors of the two methods are different and missing data exist, we use the consistency index C as the evaluation index of the accuracy. C refers to the rate of the unit number to the total number of useful observation when the forecast result and actual result are consistent.

3.2. Case study.

According to factors of electronic commerce credit, combining with Taobao existing evaluation methods and comprehensive analysis of other evaluation system and previous studies, this paper assesses C2C E-commerce credit evaluation from three aspects, including the shop information, goods information and sellers services. Shop information includes caution money, the time of setting up, punishment, top favorites, the refund rate and collection times; goods information includes description of the picture and the total transaction amount; seller services include attitude and delivery speed; credit includes credit value and credit rank (J – golden delicious; Y – silver crown).

We choose the store of "3C digital" industry as the destination, selecting the shops mainly selling watches on Taobao, collecting 260 data which will be filtered and then finally generating 242 valid data. Part of the data is shown in Table 1.

4. Control Design and Results.

4.1. Comparison of model's predictive performance with different sample sizes and missing ratios.

Simulation sample data meet the following conditions: sample data is submitted index distribution ($\lambda = 1$), the sample size is 80 and 500 and the missing ratio is 20%, 40%, 60% in the data sets.

Structural equation model: we suppose that three latent variables are all correlated with the endogenous latent variable (C2C E-commerce credit) and then establish the SEM. Fit model by adjusting MI value, other parameters default in Amos22.0 until the model is on optimal level in the corresponding T value.

BP neural network model: the 10 observed variables $(X_1, X_2, X_3, \ldots, X_{10})$ as BP neural network input neurons ($n = 10$), C2C E-commerce credit rank as BP neural network output neurons ($L = 1$), select LM optimization algorithm and hidden layer nodes, and then remain parameters default in Matlab software, and finally achieve balance of the network.

TABLE 1. Taobao shop credit evaluation index data (part)

Index / NO.	Y_1	Y_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
	Credit Value	Credit Rank	Caution Money	Time	The Cumulative Transaction	Top Favorites	Description	Attitude	Delivery Speed	Favorable Rate (%)	Refund Rate (%)	Punishment
12	4682572	3J	500	5	238	711057	4.6	4.7	4.7	98.01	1.54	0
13	1482994	2J	1000	7	210	319660	4.7	4.7	4.7	98.63	4.14	0
53	860237	1J	1000	6	5482.6	15662	4.7	4.7	4.6	98.18	8.29	0
14	683832	1J	500	8	1691	328299	4.7	4.8	4.8	99.14	1.98	0
78	318054	5y	51000	9	15960	513	4.8	4.8	4.8	99.97	7.00	0
206	290138	5y	824	8	272	182	4.7	4.8	4.7	99.10	10.51	0
11	254085	5y	1000	2.5	9504	21789	4.6	4.7	4.6	96.94	1.69	0
104	232665	5y	55000	9	2430	531	4.9	4.9	4.9	99.99	4.40	0
136	116170	4y	1000	3	680	72	4.6	4.6	4.6	98.12	3.61	0
10	112547	4y	1000	3	746704	29859	4.5	4.6	4.6	94.97	6.89	0
192	69474	3y	50000	10	41613	480	4.9	4.9	4.9	99.99	10.60	0
92	69056	3y	5000	6	16898	2017	4.8	4.8	4.7	99.99	9.78	0
159	68534	3y	7500	5	300	342	4.8	4.8	4.8	98.97	2.30	0
21	67243	3y	500	9	0	30	4.8	4.8	4.8	99.25	1.34	0
191	63711	3y	15000	3	0	180	4.6	4.7	4.7	99.30	2.30	0

Each simulation sample repeated 100 times, to get the mean and standard deviation of the consistency index C, which is shown in Table 2. The study simulates two data sets which include one sample size of 80 and the other sample size of 500.

TABLE 2. The consistency index C of the two models

Sample Size	Missing Ratio	SEM	BP	T	P
80	20%	0.779±0.021	0.785±0.028	1.248	0.201
	40%	0.767±0.016	0.771±0.023	1.187	0.243
	60%	0.748±0.024	0.746±0.017	2.425	0.018
500	20%	0.789±0.019	0.781±0.025	1.342	0.087
	40%	0.774±0.022	0.764±0.031	1.417	0.105
	60%	0.732±0.017	0.752±0.015	2.216	0.124

In the small sample data, when the missing ratio is more than 40%, C of structural equation model decreases; in the large sample data, when the missing ratio is more than 40%, C of structural equation model declines quickly. This further describes that in the same missing rate of data, the consistency index C of BP neural network model is higher than that of the structural equation model.

4.2. **Result of case study.** Use the collected data of C2C E-commerce credit evaluation to establish the SEM. Firstly, according to current theory, it is assumed that there is a significant correlation between shop information, goods information, seller's service and C2C E-commerce credit. And then we use Amos22.0 to build structural equation, inputting data and updating model to achieve optimal. The hypothesis SEM is shown in Figure 1 (e_1, e_2, e_3, ..., e_{12}, z_4 are residual variables which ensure that the model can be set up).

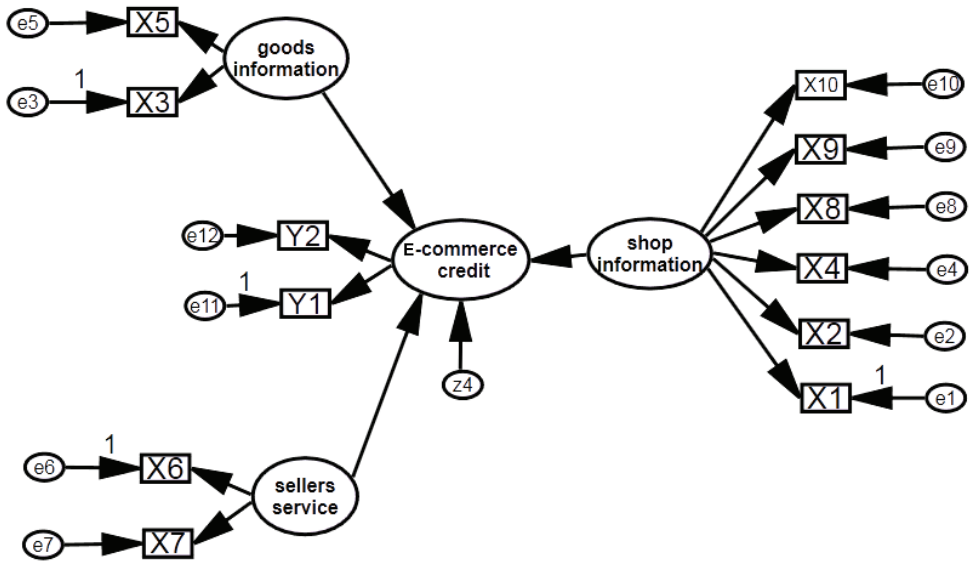

FIGURE 1. The hypothesis structural equation model

BP neural network model: we select 3-layer network architecture, use the 10 factors of C2C E-commerce credit as the network input neurons, the credit rank after clustering analysis as output neurons, the original credit rating as expectations output, based on trial and error ways to fix 5 hidden layer nodes, transfer function of hidden layer is tansig, transfer function of output layer is purelin, learning rate is 0.01, using LM optimization algorithm to train network, and applying early stop strategy in case of over-fitting. The hypothesis of BP neural network model is shown in Figure 2.

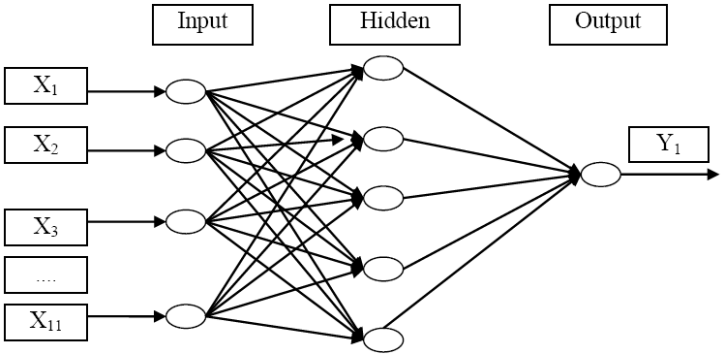

FIGURE 2. The BP neural network model

TABLE 3. The consistency index C of SEM and BP neural network

Model	Mean	Standard deviation
SEM	0.815	0.053
BP	0.837	0.047

Two models repeated 20 times to obtain the mean and the standard deviation of the consistency index C, and the results are shown in Table 3. In the case study of C2C E-commerce credit evaluation, BP neural network model's consistency index C was 0.835, which performed superior to SEM. It shows that the consistency index C of BP neural network model is higher than that of SEM.

5. **Conclusions.** For the small sample size, little missing ratio and fewer restrictions on data, BP neural network has better performance than SEM. Existing theory has proved that BP neural network's extraction of relationships has a non-specific feature; BP neural network model can overcome the traditional statistical analysis method's limitation, and does not consider the specific form of the relationship between the variables. Compared with structural equation model, BP neural network is still not used to identify potential variables and build the assumption path which is in advance of existing theories. So BP neural network is worth further popularizing in C2C E-commerce credit evaluation. The next researching direction should be the complementarity study of BP neural network and SEM in the C2C E-commerce credit evaluation.

Acknowledgment. This work is partially supported by Dalian Municipal Science and Technology Bureau Projects under Grant NO. 2014D11ZC113. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES

[1] C. Xu, Theory investigation between structural equation modeling and validity of the questionnaire structure, *Science and Technology Vision*, vol.14, no.1, p.88, 2013.
[2] X. Liu and A. Xiang, The study of customer satisfaction index of B2C E-commerce based structural equation model, *Applied Research*, vol.23, no.6, pp.60-62, 2013.
[3] Z. Jiang and W. Hu, C2C E-commerce trusting model based on structural equation, *Applied Research*, vol.28, no.2, pp.59-61, 2013.
[4] J. He, L. Zhang and S. Yang, C2C E-commerce system performance evaluation based on BP neural network, *Value Engineering*, vol.131, no.1, pp.4-6, 2008.
[5] H. Zhao and D. Wan, Comparison of structural equation modeling and artificial neural network model, *Systems Engineering-Theory Methodology Applications*, vol.7, no.3, pp.98-100, 2003.
[6] D. R. Cox and D. Oakes, *Analysis of Survival Data*, Chapman and Hall, London, 1984.

[7] W. Li, S. Huang and H. Li, Comparison of BP neural network and Cox proportional hazards model in survival analysis applications, *Journal of Zhengzhou University (Medical Sciences)*, vol.14, no.6, pp.82-84, 2014.
[8] W. Gao and L. Shi, Application of artificial neural network in epidemiology progress, *Journal of Preventive Medicine*, vol.31, no.6, pp.373-374, 2000.