Kamphol Panyagometh

INTERDISCIPLINARY APPROACH TO ECONOMICS AND SOCIOLOGY

ECONOMICS
& Sociology

IMPACT OF BASELINE POPULATION ON CREDIT SCORE'S PREDICTIVE POWER

Kamphol Panyagometh,
NIDA Business School,
Bangkok, Thailand,
E-mail: kamphol@nida.ac.th

ABSTRACT. Credit scoring involves statistical analysis performed by lenders and financial institutions to access person's creditworthiness. It utilizes statistical techniques along with debtor data such as loan application or credit bureau information to measure debtor's creditworthiness. When compared with the traditional credit evaluation process, credit scoring has shown less bias, faster speed, and consistent measurement of creditworthiness. For this reason, the National Credit Bureau of Thailand (NCB) has developed the NCB Score, based on credit behavior information collected from its members' financial institutions, which normally issue a wide variety of credit products. However, problems can arise when this NCB score is applied to a smaller bank that usually offers a few specific types of loans. As a result, the score's predictive power may deteriorate. In this paper, the impact of baseline population difference on the predictive power of a credit score was studied by separating proprietary data from NCB into two groups. One group represents those who originate personal loans in the seventh month of the study period, and the other group represents those who originate mortgage loans in the seventh month of the study period. The credit score model of each group was developed and their predictive power was compared when used with the same baseline population and a different baseline population to monitor the change in model predictive power

Introduction

Traditionally, credit evaluation process relies on underwriter experience to judge between good and bad applicants and whether the credit should be granted or not. However, the process is subjective depending on the bias of underwriter and person-to-person interpretation, resulting in inconsistent decision-making, in which case two underwriters might come to different judgements for the same applicant. Due to the improvement of computing technologies, statistical techniques together with applicant data have been used to determine credit scores that measure creditworthiness and determine the likelihood of credit payment. A credit score helps automate the credit valuation process in a fast, accurate, and consistent way,

thus making credit opportunities more accessible to applicants (Abdou & Pointon, 2011; Hand & Henley, 1997).

Since any credit score is calculated based on data, the credit score model is said to be data-driven, which means every step in calculating the credit score model must rely on data. For example, if data used to create the credit score are the data collected from loan application information such as gender, sex, age, educational level, occupation, annual income, source of income, and income-to-debt ratio, such a credit score is called an "application score." On the other hand, if the data used to create a credit score are credit bureau data such as payment history, loan amount, loan use, and loan amount outstanding, such credit score is called a "bureau score" (Vojtek & Koèenda, 2006)

In Thailand, NCB has developed the NCB Score based on a large pool of debtor behaviors for every loan to obtain a single score to represent individual creditworthiness. Since the NCB Score is easy to access, it has widely been adopted and became traditional for the institutions without their internal model and as a benchmark for the banks that already have internal models of their own. However, since the NCB Score was developed from various credit portfolios, the adoption of the score by small banks that have different baseline populations may raise concerns about the predictive power of NCB Score for a small bank portfolio.

In order to study the impact of different baseline populations on a credit score's predictive power, two credit score models were constructed using two data sets as a representative of two baseline populations, and each model was used to predict the probability of default of one another population used to develop the model.

1. Literature review

Mester (1997) stated that credit scoring is an analysis of credit risk for loan approval by using past information and statistical approaches to create a model or scorecard in assessing loan repayment. The borrower's characteristics, such as salary, total remaining debt, or years of work service, are key information aspects to forecast the efficiency of loan repayment. Although credit scoring cannot be accurately predicted in all cases, it offers a way to identify the related risk of each borrower group. Avery, Brevoort, and Canner (2009) also extended the study by stating that the first important aspect of credit scoring is the accuracy and fairness of the report and scoring system, which is in accordance with the study of Chuang and Huang (2011) and Sengupta, Rajdeep, Bhardwaj, and Geetesh (2015). In the past, most factors used in credit scoring included structured data, which includes the debt repayment information in the past and the demographic features of the borrower. Bernerth (2012) conducted a study on the relationship of demographic variables and individual clients by using demographic data and credit data gained from a total of 142 employees. It was found that minority status was negatively correlated, while age and education were positively correlated with credit score. However, the previous research by Capon (1982) showed that certain economic factors may be useful for the analysis of credit scoring, but they are not always applicable, such as postcode or revenue, because they are less statistically related to other variables. In addition, research by Avery, Calem, Canner, and Glenn (2004) stated that although the borrower's past data is very useful in but the loan approval decision, the credit scoring model that only considers data from the past fails to consider the effective data for the time being.

Recently, there have been numerous studies on credit scoring, such as the study conducted by Hooman et.al (2016). This study was performed by applying data mining to credit scoring using several approaches, including discriminant analysis, logistic regression, K-nearest neighbor, Bayesian classifier, decision tree, neural network, survival analysis, fuzzy rule-based system, support vector machine, and hybrid methods, with an aim to find the best

analytical approach. However, Khashei and Mirahmadi (2015) stated that there is no widely accepted model that is accurate and able to explain all circumstances.

Therefore, unstructured data has been developed to create a model. Ntwiga and Weke (2016) stated that the traditional consumer credit score is limited because of the need for financial history of the borrower. In the absence of financial background, it is difficult to perform credit analysis on the borrower, and as a result, the bank may lose an opportunity to make profit. Notwithstanding, with advanced technology, there is a concept to apply the social data to credit analysis. They found that the use of social data helps in providing more useful information than with the use of a traditional credit scoring technique.

This was in accordance with the statement of Masyutin (2015), which specified that social networks have kept a record of user behaviors that can be used in identifying the borrower's credit information. He discovered that the information from social networks alone can offer results that were accurate at an acceptable level. He also discovered that the use of information from social networks can identify the borrower who intended from the beginning not to repay debt in addition to the borrower with potential default.

There was also a study conducted on the past relationship. Chakravarty and Scott (1999) studied the relationship between the households and the creditors which affects the probability of credit-rationing. He found that the relationship duration and the number of activities between a family and a potential lender significantly decrease the probability of being credit-rationed due to the lower risk of failure to repay the principal amount.

With regard to the popular online P2P lending market, Li, Lin, Qui, Safi, and Xiao (2015) conducted a study on information from PPDai.com, a national online P2P platform leader, to examine the risk of default. The researchers had created a credit scoring model from the integration of social network information based on the decision tree model. The test results showed that the most important factors in forecasting default were loan information, social media information, and credit information.

In addition, Zhanga, Jiaa, Diaoa, Haia, and Lia (2016) conducted a further study on the effects of the multidimensional friendship network and economic outcomes on the P2P market, which is where people can directly borrow and lend small amounts of money. This research used empirical analysis based on transaction data and the friendship network from PPDai.com. The researchers found that there is a significant correlation between the number of friends on the network and the loan. In other words, the borrower who had a good network of friends will be granted a loan, and a lower rate of interest will be charged. Also, the result was more significant if the borrower had a larger network of friends in the friendship hierarchy. In other words, greater quality of the friend network will lead the lender to believe that the borrower is more reliable. Furthermore, the research found that the quality of friends is more important than the number of friends on the network when making considerations in granting loans.

2. Data and Methodology

To study the impact of the baseline population on a credit score's predictive power, the proprietary individual loan records from April 2015 to September 2016 (18 months) were provided by NCB, and each record contains identification information (identification code, account status, account type, term frequency, etc.) and loan performance information (payment history, credit limit, outstanding balance, days past due, etc.).

Two data sets were constructed (mortgage loan and personal loan). The loan records of individuals who originated mortgage loans in the 7th month of the study period are referred to as the mortgage data set (16,692 individuals and 95,463 loan records with 3,468 bad loans and 91,995 good loans), and loan records of individuals who originated personal loans in the 7th

month of the study period (156,451 individuals 787,727 loan records with 60,531 bad loans and 727,196 good loans) are referred to as the personal loan data set.

The data set then undergoes a data transformation to create a more predictive variable with meaningful behavior. The transformation includes utilization, momentum, standard deviation, delay payment within specific period, and number of delay payment. In this study, the good and bad debtors are defined as those who do not miss payments for more than 90 days within next 12 months and those who missed payments for more than 90 days within next 12 months, respectively.

The mortgage credit score and personal loan credit scores were developed using the mortgage data set and personal loan data set, respectively. The model developed in this study was selected to mimic the NCB score, that is, different types of loans should have different credit score models. However, since the NCB score model is confidential and not publishable, a two-stage model was developed instead (Steinberg & Cardell, 1998; Lapczynski, 2016). For the first stage, the classification trees are constructed for each type of loan and assigned a terminal node with a dummy loan score from 1 to 7 (with the lowest probability of default assigned a value of 1 and the highest probability of default assigned a value of 7) (Figure 1).

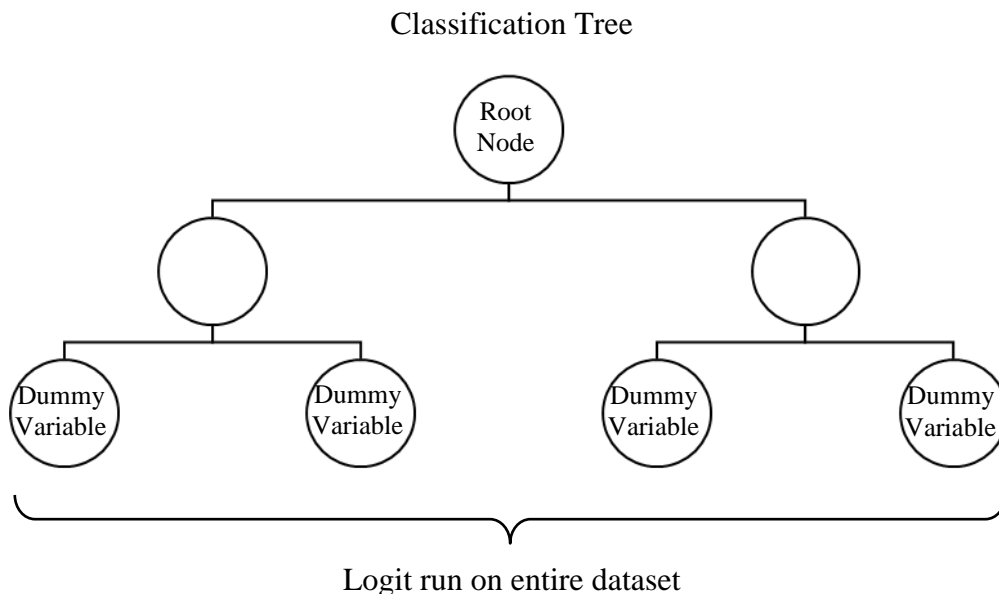Classification Tree

Logit run on entire dataset

Figure 1. 2 Stage models used in developing mortgage and personal loan credit scores.

Due to imbalanced nature of the data set, most of the classification trees do not grow. To overcome such an issue, over-sampling techniques were used to adjust imbalanced good-to-bad loan ratio and the classification tree branching factor. Such as complexity parameter and depth of the tree, was varied to find the tree with a sufficient leaf node for the dummy loan score.

The performance of each classification tree is validated using the area under receiver operating characteristics curve (AUC). The receiver operating characteristics (ROC) curve is a plot between true positive on the Y-axis and true negative on the X-axis used to measure a classifier's performance. When using the ROC curve, it is often hard to provide a clear distinction between classifiers; therefore, the AUC is used instead to provide one summary number that indicates classifier performance or predictive power. AUC varies between 0 to +1, and the larger AUC, the better the classifier predictive power (Huang & Ling 2005).

The rules from the classification trees were used to transform the data set by assigning every loan record with a dummy loan score and reducing the data set from a per loan record to per individual by summarizing the number of loan scores each individual had.

Kamphol Panyagometh

INTERDISCIPLINARY APPROACH TO ECONOMICS AND SOCIOLOGY

In the second stage, the data sets were transformed by generating the probability of the loan score each individual had from number of loans and the number of loan scores. Then the logistic regression model was constructed using the probability of the loan score as a predictor variable and validated by AUC.

$$\log\left(\frac{p_i}{1 - p_i}\right) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \cdots + x_n\beta_n$$

Where: p_i is the probability of default

β_n is the regression coefficient for the probability of loan score n

x_n is the probability of the loan score n $\left(\frac{number\ of\ loan\ score\ n}{total\ number\ of\ loan\ score}\right)$

Next the impact of the difference in baseline population to the credit score's predictive power was measured by feeding the mortgage model developed from the mortgage data set with the personal loan data set and measuring the predictive power using AUC, and vice versa.

3. Results and Discussion

From the variation of over-sampling and classification tree branching factors, the classification tree developed using the mortgage data set was shown to grow sufficiently for the dummy loan score using an over-sampling ratio of 80:20, complexity parameter of 0.005, and maximum tree depth of 5. On the other hand, the classification tree using personal loan data was difficult to grow compared to the tree developed from the mortgage data set. The over-sampling ratio of 80:20, complexity parameter of 0.001, and maximum depth of tree 8 were used to produce the tree sufficient for loan scoring. Table 1 summarized the over-sampling and classification tree branching factor used in the model development.

Table 1. Over-sampling and classification tree branching factor used in model development

	Mortgage Model	Personal Loan model
Good to Bad loan ratio	80:20	80:20
Complexity Parameter	0.005	0.001
Maximum depth of Tree	5	8

Table 2 and Table 3 reported the AUC of the classification tree developed using the mortgage data set and personal loan data set respectively. From Table 2 and Table 3, the classification tree developed using the mortgage data set was shown to have higher predictive power (higher AUC) for overdraft, personal loan, mortgage, credit card, and auto hire purchase. The classification tree developed using the personal loan data set was shown to have higher predictive power (higher AUC) for hire purchase, commercial loan, and other loan.

Table 2. AUC of each classification tree developed in the mortgage model

Type of Loan	AUC
Overdraft	0.7245
Personal Loan	0.6526
Mortgage	0.7672
Hire Purchase	0.529
Credit Card	0.6913
Auto Hire Purchase	0.7233
Commercial Loan	0.6489
Other Loan	0.5454

Table 3. AUC of each classification tree developed in the personal loan model

Type of Loan	AUC
Overdraft	0.6907
Personal Loan	0.5911
Mortgage	0.6232
Hire Purchase	0.5684
Credit Card	0.5762
Auto Hire Purchase	0.6079
Commercial Loan	0.6808
Other Loan	0.56

Table 4. AUC of each credit score model used to predict the probability of default with the different data sets

	Mortgage Model	Personal Loan Model
Mortgage data set	0.7852	0.6766
Personal Loan data set	0.7017	0.6767

Table 4 showed that the predictive power from the mortgage model dropped by 12% when used to predict the probability of default of the personal loan data set. This is due to the fact that the mortgage data set used to develop the mortgage model was comparatively smaller when compared to the personal Loan data set, hence the mortgage model might not account for the additional feature of data persisting in the personal loan data set. Therefore, the mortgage model was less of a fit with a personal loan data set and resulted in a significant drop in predictive power when used to predict the probability of default of personal loan data.

On the other hand, the personal loan model predictive power did not have a significant impact when used to predict the probability of default of mortgage data. Since the personal loan data set was larger than the mortgage data set, and the model reflected all behavior persisting in the mortgage data set leading to a model with insignificant impact on predictive power when used to predict probability of default of mortgage data set.

Even though the Mortgage model was used to predict probability of default of the personal loan data set, the predictive power of the mortgage model was higher than the personal loan model's capability to predict the probability of default of the personal loan data set. This might be related to the following reasons:

1. When constructing the mortgage model, a higher complexity parameter and lower maximum depth were selected compared with the personal loan model.
2. The delinquent nature for the specific loan is that the individual is more likely to miss payment on a personal loan in comparison to a mortgage loan when they faced financial problems resulting in unclear classification criteria and poor predictive power.

The nature of the loan is that the mortgage loan usually consists of an installment loan while the personal loan may have both an installment loan and revolving loan. This leads to a possible difference in debtor behavior which in turn makes the personal loan data set difficult to classify in comparison to the mortgage data set.

Conclusion

The NCB Score has been wildly adopted to measure creditworthiness of a debtor; however, the NCB Score's predictive power might be affected by baseline population for a small bank portfolio, which is more specific and less diverse than the NCB database. The impact

of the baseline population used in credit score development has been studies by developing a mortgage score model from a mortgage data set and a personal loan score model from a personal loan data set using two-stage models. The first stage used a classification tree to assign a dummy loan score for each loan and summarized each loan score with an individual number of loans. In the second stage, logistic regression was used to predict probability of non-payment more than 90 days within next 12 months.

According to the study, when the mortgage model was used to predict probability of default of personal loan data set, the predictive power of the model dropped by 12% which was very significant. Although when the personal loan model was used to predict probability of default of mortgage data set, the predictive power did not change significantly.

The study showed that the model developed from a larger data set (i.e. personal loan data set) was more robust and less sensitive when compared with the model developed from a smaller data set. However, being more robust and less sensitive did not imply that the model had higher predictive power compared to the model developed from a smaller data set. Hence, when developing a credit score, the larger amount of data available did not imply a better credit score model. The nature of the use of data also impacted credit score significantly.

Since the score was developed from various credit portfolios collected from bank members, the predictive power of the NCB Score did not show significant impact when applied to a small bank portfolio unless such portfolio had not been covered in the NCB database.

Acknowledgement

References

Acemoglu, D., Johnson, S., Robinson. I., & Thaicharoen, Y. (2004). Institutional Causes. Macroeconomic Simptoms: Volatility. Crises and Growth. *Journal of Monetary Economics*, 50, 49-123.

Abdou, H. A., & Pointon, J. (2011). Credit scoring, statistical techniques and evaluation criteria: A review of the literature. *Intelligent Systems in Accounting, Finance and Management*, 18(2-3), 59-88.

Ambrus, A., Mobius, M., Szeidl, A. (2014). Consumption risk-sharing in social networks. *The American Economic Review*, 104(1), 149-182.

Avery, R. B., Brevoort, K. P., & Canner, G. B. (2009). Credit scoring and its effects on the availability and affordability of credit. *The Journal of Consumer Affairs, 43*(3), 516-537.

Berger, A. N., Cowan, A. M., Frame, W. S. (2011). The surprising use of credit scoring in small business lending by community banks and the attendant effects on credit availability, risk, and profitability. *Journal of Financial Services Research, 39*(1-2), 1-17.

Bernerth, J. B. (2012). Demographic variables and credit scores: An empirical study of a controversial selection tool. *International Journal of Selection and Assessment*, 20(2), 242-246.

Canbolat, M. A., & Sevim, S. (2016). A novel software model for credit requests rating: Scoring model. International Journal of Innovation, *Management and Technology, 7*(1), 22-26.

Capon, N. (1982). Credit scoring systems: A critical analysis. *Journal of Marketing, 46*(2), 82-91.

Chakravarty, S., & Scott, J. S. (1999). Relationships and rationing in consumer loans. *The Journal of Business*, 72(4), 523-544.

Chuang, C., & Huang, S. (2011). A hybrid neural network approach for credit scoring. *Expert Systems, 28*(2), 185-196.

Crook, J. N., Hamilton, R., Thomas, L. C. (1992). A comparison of a credit scoring model with a credit performance model. *The Service Industries Journal, 12*(4), 558-579.

DeYoung, R., Glennon, D., & Nigro, P. (2008). Borrower-lender distance, credit scoring, and loan performance: Evidence from informational-opaque small business borrowers. *Journal of Financial Intermediation, 17*(1), 113-143.

Einav, L., Jenkins, M., & Levin, J. (2013). The impact of credit scoring on consumer lending. *The Rand Journal of Economics, 44*(2), 249-274.

Hand, D. J., & Henley, W.E. (1997). Statistical classification methods in consumer credit scoring: a review. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 160(3), 523-541.

Hooman, A., Marthandan, G., Yusoff, W. F. W., Omid, M., & Karamizadeh, S. (2016). Statistical Mining Methods in Credit Scoring. *The Journal of Developing Areas*, 50(5), 371-381.

Huang, J., & Ling, C.X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3), 299-310.

Łapczyński, M. (2016). The Use of Hybrid Predictive C&RT-Logit Logit Models in Analytical CRM. *Analysis of Large and Complex Data*, Springer, 311-320.

Khashei, M., & Mirahmadi, A. (2015). A soft intelligent risk evaluation model for credit scoring classification. *International Journal of Financial Studies, 3*(3), 411-422.

Kunvipusilkul, D. (2008). Credit information in Thailand. Session 1: An overview of challenges related to measuring financial innovations, 26: 49.

Masyutin A.A. (2015). Credit scoring based on social network data. *Business Informatics*, 3(33), 15–23.

Mester, L. J. (1997). What's the point of credit scoring? *Business Review* - Federal Reserve Bank of Philadelphia, 3-16.

Ntwiga, D.R., & Weke, P. (2016). Consumer lending Using Social Media Data. *International Journal of Scientific Research and Innovative Technology*, 3(2), 1-8.

Sengupta, R., & Bhardwaj, G. (2015). Credit scoring and loan default. *International Review of Finance, 15*(2), 139-167.

Sohn, S. Y., & Kim, Y. S. (2013). Behavioral credit scoring model for technology-based firms that considers uncertain financial ratios obtained from relationship banking. *Small Business Economics*, 41(4), 931-943.

Spader, J. S. (2010). Beyond disparate impact: Risk-based pricing and disparity in consumer credit history scores. *Review of Black Political Economy, 37*(2), 61-78.

Steenackers, A., & Goovaerts, M. J. (1989). A credit scoring model for personal loans. *Insurance, Mathematics & Economics, 8*(1), 31-34.

Steinberg, D., & Cardell, N. S. (1998). The hybrid CART-Logit model in classification and data mining. Salford Systems White Paper.

Vojtek M., & Koèenda, E. (2006). Credit-scoring methods. *Czech Journal of Economics and Finance (Finance a uver)*, 56(3-4), 152-167.

Wei, Y., Yildirim, P., Van den Bulte, C., & Dellarocas, C. (2015). Credit scoring with social network data. *Marketing Science*, 35(2), 234-258.

Zhang, Y., Jia, H., Diao, Y., Hai, M., & Li, H. (2016). Research on credit scoring by fusing social media information in online peer-to-peer lending. *Procedia Computer Science*, *91*, 168-174.