Proceedings of the 7th International Conference on Economic Management and Green Development
DOI: 10.54254/2754-1169/42/20232097

The Predictive Power of Credit Scores: Examining Default Probability in Taiwanese Credit Card Clients

Yaoxin Xiao[1,a,*]

[1]*St. George Campus, University of Toronto, Toronto, ON, M5S 1A1, Canada*
a. yaoxin.xiao@mail.utoronto.ca
**corresponding author*

Abstract: The concept of a scorecard originated from the need to establish a standardized and objective approach to evaluate credit applicants. Various techniques have been utilized to build scoring model. This research chooses Logistic regression to construct a scorecard using SPSS modeler. In this way, the study enhances the understanding of the relationship between credit scores and default behavior. By using a scorecard constructed through logistic regression, the study provides a comprehensive and interpretable model for evaluating creditworthiness. The study also employs linear probability models (LPM), logit, and probit models to assess the predictive power of credit scores on default probability. By utilizing these statistical techniques, the research presents robust empirical evidence on the significance of credit scores in predicting default behavior. Moreover, the research paper systematically analyzes prediction effects with and without control variables. By incorporating control variables such as demographic characteristics, the study adds depth to the understanding of scoring models. Overall, the findings provide valuable guidance for credit risk assessment practices and contribute to the ongoing development of effective credit evaluation frameworks.

Keywords: credit scoring, predictions, statistical learning, machine learning

1. Introduction

The relationship between credit scores and default behavior has been studied extensively in the field of risk management. According to Gjini et al., credit risk has become the largest risk faced by financial institutions [1]. Understanding the relationship between credit scores and default behavior is essential for effective risk assessment and responsible lending practices. It is crucial to establish a method for distinguishing between reliable and unreliable candidates. To address this challenge, financial institutions have proactively developed credit scoring systems as a viable solution. To enhance the accuracy and effectiveness of scoring systems, a growing number of classification methods have been developed [2].

In 1996, Ripley used Neural Networks (NN) to build credit scoring systems which simulated human brains in solving binary classification problems [3]. Later, Feng et al. proposed a new credit scoring system using Support vector machine (SVM) [4]. Berkson introduced Logistic Regression (LG) in 1944 as a conventional classification technique for constructing scoring models [5]. Over the years, logistic regression has been widely employed and extensively practiced. Additionally, decision

DOI: 10.54254/2754-1169/42/20232097

trees, linear regression, Bayesian networks, and various other techniques have also been utilized in the construction of scoring models [6-8].

Based on the comprehensive review on binary classification techniques for credit scoring provided by Bücker et al., this study chooses Logistic regression model as the key method to build a scorecard. The primary justification for this choice is that logistic regression models are not only user-friendly but also highly interpretable. Logistic regression offers a transparent interpretation of the relationship between the input variables and the output, specifically the probability of a particular outcome. The coefficient associated with each variable enables us to comprehensively comprehend the influence and importance of each variable within the scoring model [9].

This research paper aims to investigate the relationship between credit scores and default behavior. To accomplish this objective, the study adopts a two-step approach. In Section 2, credit scores are calculated using SPSS Modeler, a reliable tool for credit scoring that utilizes statistical techniques and machine learning algorithms. These credit scores serve as quantitative indicators of creditworthiness for each individual in the dataset. In Section 3, the correlation between credit scores and default behavior was examined using STATA. Linear probability models (LPM), Logit and Probit models were employed to assess the predictive power of credit scores.

When testing predictability in STATA, two sets of models were analyzed: the first set explored the correlation between credit scores and default without incorporating any control variables, while the second set incorporated additional individual characteristics as control variables. These individual characteristics included demographic factors such as age, education, and sex. This approach allowed for a comprehensive analysis, considering potential confounding factors that may influence the relationship between credit scores and default behavior.

The primary objective of this research is to test whether user scores derived from a scorecard calculation method have predictive power in determining whether individuals will default on their credit card payments. The results obtained from the analysis indicated that credit scores were indeed significant in predicting default behavior.

2. The Scoring Model

2.1. Data Exploration

The dataset obtained from Kaggle will be adopted for this study. The dataset used in this study encompasses a wide range of variables, including payment history, demographic factors, bill statements, and credit information of 30,000 credit card clients in Taiwan from April 2005 to September 2005. This rich dataset serves as a valuable resource for investigating the intricate connection between credit scores and default behavior among Taiwanese clients. There is no missing data in the entire dataset, and all variables are numeric values. In building the logistic regression model, the “DEFAULT” variable is set to be the target variable (0 represents non-default, 1 represents default), and the rest of the factors are input variables. Summary Statistics are shown in Table 1.

Table 1: Summary statistics.

Variables	Obs	Min	Mean	Max	SD
SEX (1=male, 2=female)	30000	1	1.6037	2	0.4891
EDU	30000	0	1.8531	6	0.7903
MARRIAGE	30000	0	1.5518	3	0.5219
AGE	30000	21	35.4855	79	9.2179

Table 1: (continued).

Repay_0	30000	-2	-0.0167	8	1.1238
Repay_2	30000	-2	-0.1337	8	1.1971
Repay_3	30000	-2	-0.1662	8	1.1968
Repay_4	30000	-2	-0.2206	8	1.1691
Repay_5	30000	-2	-0.2662	8	1.1331
Repay_6	30000	-2	-0.2911	8	1.1499
LIMIT_BAL	30000	10000	167484.3226	1000000	129747.6615
DEFAULT	30000	0	0.2212	1	0.4150

2.2. Build the Scorecard Model

2.2.1. Binning

Generating a scoring system basically involves three steps [10]. First, relevant input variables should be regrouped using weight of evidence (WOE). The risk control modeling process often necessitates the binning of variables, primarily to transform continuous variables into categorical variables and allow for suitable combinations of these categories. The main goal of binning is to enhance the overall stability of the model. By grouping similar values together and creating distinct categories, the variability in each category can be reduced, leading to a more robust and reliable model. Binning helps in managing outliers, reducing noise, and capturing nonlinear relationships, thereby improving the performance and interpretability of the model. For continuous variables, there is a special binning node in SPSS Modeler, which contains the optimal binning method [11]. To use this method, a discrete target variable needs to be set. Once the target variable is set, the node bins the specified continuous input variables based on the distribution of that target variable, and then optionally exports the results of binning the continuous variables to generate new variables.

After dividing each input variable into new bins, WOE and information value (IV) needs to be calculated. The formula for WOE is:

$$\text{goods} = (\text{good clients in the bin/total good clients in the sample}) \tag{1}$$

$$\text{bads} = (\text{bad clients in the bin/total bad clients in the sample}) \tag{2}$$

$$\text{WOE} = \ln(\frac{\text{goods}}{\text{bads}}). \tag{3}$$

Further, the IV value can be calculated based on the WOE value, and the IV value is calculated by the formula:

$$\text{iv} = \sum_{i=1}^{n}(\text{goods}_i - \text{bads}_i) \times \text{WOE}_i \tag{4}$$

Information Value (IV) plays a significant role in binary classification models, particularly in the context of credit scoring or risk modeling. IV is a metric used to assess the predictive power of a given variable in distinguishing between two classes (e.g., good credit vs. bad credit).

The main role of IV is to evaluate the strength of the relationship between an independent variable and the dependent variable (target variable) in a binary classification model [12]. It quantifies the discriminatory power of each variable by measuring the variable's ability to separate the positive and negative classes. In practical terms, the higher the IV value for a variable, the stronger its predictive power. Variables with high IV values are considered more informative and useful in predicting the

DOI: 10.54254/2754-1169/42/20232097

outcome of interest. After calculating the bins of each variable and the WOE and IV information values corresponding to the bins, the study excluded variables with total iv values greater than 0.5 and total iv values less than 0.03. Eventually, the logistic regression model incorporated the variables Repay_3, Repay_4, Repay_5, Repay_6, EDU, and LIMIT_BAL.

2.2.2. Build Logistic Regression Models

Typically, input variables in credit score card construction predominantly consist of continuous variables. In cases where the input variables are discrete, it is common to employ dummy variables to convert them into continuous variables before further processing. However, when developing a credit score card, all variables have already been transformed into discrete variables through binning. As an alternative approach, the WOE values corresponding to each bin of a variable are commonly used as input variables in Logistic regression [13]. This methodology ensures that the distinctions between different bins are adequately captured while preserving the trend of each variable's distribution with respect to the target variable. By utilizing WOE values, the credit score card construction process can effectively incorporate the information present in the discrete variables without sacrificing crucial insights. This approach also maintains the integrity of the variable relationships. In SPSS modeler, new bins and their corresponding WOE values were merged into the original dataset. Next, the dataset was separated into training and testing groups, then a stepwise logistic regression model was adopted. Formula 5 shows the resulting coefficients of selected input variables regarding their WOE. Finally, the study summarized variables that went into the logistic regression model as well as each bin's WOE, coefficients, and constants into excel (Table 2).

$$\text{Default} = \text{repay6WOE} \times -0.05344 + \text{repay5WOE} \times -0.07249 + \text{repay4WOE} \times -0.04414 + \text{repay3WOE} \times -0{,}121 + \text{eduWOE} \times -0{,}05521 + \text{creditWOE} \times -0.0838 + 0.2454 \quad (5)$$

Table 2: Logistic regression summary.

Variables	Binning	WOE	Coef.	_Cons
Repay_6	1	0.213	-0.0534	0.2454
Repay_6	2	-1.285	-0.0534	0.2454
Repay_6	3	-1.840	-0.0534	0.2454
Repay_6	4	-1.802	-0.0534	0.2454
Repay_6	5	-1.413	-0.0534	0.2454
Repay_6	6	-2.288	-0.0534	0.2454
Repay_6	7	-2.868	-0.0534	0.2454
Repay_5	1	0.228	-0.0725	0.2454
Repay_5	2	-1.427	-0.0725	0.2454
Repay_5	3	-1.812	-0.0725	0.2454
Repay_5	4	-1.615	-0.0725	0.2454
Repay_5	5	-1.694	-0.0725	0.2454
Repay_5	6	-2.848	-0.0725	0.2454
Repay_5	7	-2.357	-0.0725	0.2454
Repay_4	0	0.261	-0.0441	0.2454

Table 2: (continued).

Repay_4	1	-1.259	-0.0441	0.2454
Repay_4	2	-1.352	-0.0441	0.2454
Repay_4	3	-1.711	-0.0441	0.2454
Repay_4	4	-1.952	-0.0441	0.2454
Repay_4	5	-1.316	-0.0441	0.2454
Repay_4	6	-0.853	-0.0441	0.2454
Repay_4	7	-2.827	-0.0441	0.2454
Repay_4	8	-1.259	-0.0441	0.2454
Repay_3	0	0.313	-0.1210	0.2454
Repay_3	1	-0.160	-0.1210	0.2454
Repay_3	2	-1.321	-0.1210	0.2454
Repay_3	3	-1.561	-0.1210	0.2454
Repay_3	4	-1.577	-0.1210	0.2454
Repay_3	5	-1.546	-0.1210	0.2454
Repay_3	6	-1.701	-0.1210	0.2454
Repay_3	7	-2.740	-0.1210	0.2454
Repay_3	8	-1.952	-0.1210	0.2454
credit	1	-0.686	-0.0838	0.2454
credit	2	-0.171	-0.0838	0.2454
credit	3	0.362	-0.0838	0.2454
credit	4	0.746	-0.0838	0.2454
edu	1	0.178	-0.0552	0.2454
edu	2	-0.091	-0.0552	0.2454
edu	3	-0.168	-0.0552	0.2454
edu	4	1.549	-0.0552	0.2454
edu	5	1.419	-0.0552	0.2454
edu	6	0.423	-0.0552	0.2454

2.2.3. Converting Logistic Regression Coefficients into Scores.

A general equation to represent the values taken for credit scores is as follows [14]:

$$\text{odds} = \frac{\text{goods}}{\text{bads}} \tag{6}$$

$$\text{Score} = \text{offset} + \text{factor} \times \ln(\text{odds}) \tag{7}$$

$$\text{Score} + \text{pdo} = \text{offset} + \text{factor} \times \ln(2 \times \text{odds}) \tag{8}$$

Where pdo (points to double the odds) indicates the score that needs to be increased in order to double the odds. Since in Logistic Regression,

DOI: 10.54254/2754-1169/42/20232097

$$\text{logit}(p) = \ln(\frac{p}{1-p}) = \ln(\text{odds}) = \beta_0 + \beta_1 x_1 + \ldots + \beta_k x_k \quad (9)$$

Corresponding rating value for each variable can be computed by the following equation:

$$\text{Score} = \frac{\text{offset}}{n} - \text{factor} \times (\frac{\beta_0}{n} + \beta_i \times \text{WOE}_i). \quad (10)$$

This study took the odds to be 30:1 for a basic value of 500 and 50 for pdo, therefore:

$$\text{factor} = \frac{50}{\ln(2)} \quad (11)$$

$$\text{offset} = 500 - \frac{50}{\ln(2)} \times \ln(30) \quad (12)$$

In SPSS modeler, a new column “scorecard” (Table 3) was created based Table 2. This new column was generated through equation 10, and it contains the corresponding score for each bin. After producing the scorecard, each client’s total credit score is generated by adding up the points across all variables for the client.

Table 3: Scorecard summary.

Variables	Binning	Scorecard	Variables	Binning	Scorecard
Repay_6	1	48	Repay_4	0	48
Repay_6	2	42	Repay_4	1	43
Repay_6	3	40	Repay_4	2	43
Repay_6	4	40	Repay_4	3	42
Repay_6	5	42	Repay_4	4	41
Repay_6	6	39	Repay_4	5	43
Repay_6	7	36	Repay_4	6	45
Repay_5	1	49	Repay_4	7	38
Repay_5	2	40	Repay_4	8	43
Repay_5	3	38	Repay_3	0	50
Repay_5	4	39	Repay_3	1	46
Repay_5	5	39	Repay_3	2	36
Repay_5	6	32	Repay_3	3	34
Repay_5	7	35	Repay_3	4	34
credit	1	43	Repay_3	5	34
credit	2	46	Repay_3	6	33
credit	3	50	Repay_3	7	23
credit	4	52	Repay_3	8	30
edu	1	48			
edu	2	47			
edu	3	47			
edu	4	54			
edu	5	53			
edu	6	49			

DOI: 10.54254/2754-1169/42/20232097

3. Estimation Outcomes

Once the credit scores for each client have been computed using the scoring model developed in Section 2, the next step in this study is to investigate the correlation between credit scores and default behavior. To achieve this, the analysis employs various statistical models, including the linear probability model, Logit model, and Probit model in STATA. These models offer a comprehensive framework to assess the relationship between credit scores and the likelihood of default. It is worth noting that the coefficients output from a logit or probit regression cannot be interpreted directly [15]. In this study, the focus is on interpreting the marginal effects of the regressors. These effects reflect the change in the probability of the outcome variable when modifying the value of a single regressor while keeping all other regressors constant at specific values. To explain the coefficients using probability, it is necessary to calculate the average partial effect (APE). The APE calculated the average change in the probability of the outcome associated with a one-unit change in the independent variable, ceteris paribus [16].

3.1. Baseline Regression Results

In this model, no control variables were involved. The study used only score as independent variables and default as dependent variables. The regression coefficients demonstrate marginal effects. Based on the estimation results presented in Table 4, it can be observed that a one-point increase in the credit score leads to a statistically significant decrease of 1.36 percentage points in default probability, holding all other variables constant based on the 1% significance level. A p-value of "0.000" indicates that the p-value is very small and essentially zero. This suggests that the score is highly statistically significant in predicting default. An R-squared value of 0.1212 in LPM indicates that approximately 12.12% of the variation in the probability of default can be explained by the credit score alone. The logit model yielded a Pseudo R-squared of 0.0988, while the probit model resulted in a Pseudo R-squared of 0.0995. Although these two values are relatively close, it is important to note that Pseudo R-squared cannot be directly compared between logit and probit models.

Table 4: Baseline model results.

Variables	(1)		(2)		(3)	
	LPM		Logit (APE)		Probit (APE)	
	Coef.	p>|t|	Coef.	p>|z|	Coef.	p>|z|
Score	-0.0135	0.000	-0.0101	0.000	-0.0107	0.000
Constant	10.8849	0.000				
R-sq	0.1212					
Pseudo R-sq			0.0988		0.0995	

3.2. Marginal Effects of Regressors with Control Variables

Next, to improve model accuracy and address confounding factors in the models, the study introduced several control variables. Before running the three models in STATA, the study required some pre-processing of the data. First, the square of age should be included as a covariate to capture potential non-linear relationships between age and the probability of the outcome. Second, discrete variables need to be transformed into dummy variables so that models can capture group-specific effects.

Table 5: Marginal effect.

Variables	(1)		(2)		(3)	
	LPM		Logit (APE)		Probit (APE)	
	Coef.	p>|t|	Coef.	p>|z|	Coef.	p>|z|
Score	-0.0136	0.000	-0.0101	0.000	-0.0106	0.000
Female=1	-0.0178	0.000	-0.0190	0.000	-0.0188	0.000
Married=1	0.0289	0.000	0.0283	0.000	0.0280	0.000
Age	-0.0018	0.320	-0.0034	0.043	-0.0035	0.053
Age-sq	0.0003	0.164	0.0005	0.019	0.0004	0.022
Graduate=1	0	(omitted)	0	(omitted)	0	(omitted)
University=1	0.1598	0.000	0.0188	0.744	0.0187	0.739
High School=1	0.1471	0.000	0.0143	0.803	0.0134	0.812
Others=1	0.1428	0.000	0.0120	0.834	0.0113	0.842
Unknown=1	0.1549	0.000	-0.0800	0.329	-0.0596	0.422
Unknown=1	0.1207	0.000	-0.0977	0.155	-0.0790	0.221
Unknown=1	0.1394	0.007	0	(omitted)	0	(omitted)
Constant	10.8097	0.000				
R-sq	0.1223					
Pseudo R-sq			0.1014		0.1020	

According to the new results, control variables such as sex and marriage have small p-value, suggesting that these control variables might have a strong impact on the probability of the dependent variable in the LPM. Apart from that, credit scores still demonstrate significant power in predicting default probability. Take the APE of probit model as an example (Table 5), on average, across all the observations in the sample, an additional point in credit score decreases the probability of default by 1.06 percentage points, ceteris paribus. The significance of the credit score variable even after controlling for demographic characteristics implies that the additional factors considered in the score calculation play a crucial role in determining default behavior. The LPM with control variables has an R-squared of 0.1236. The difference in R-squared (0.1236 - 0.1212 = 0.0024) suggests that the additional control variables in the LPM model contribute slightly to explaining the variation in the default probability beyond what can be explained by the credit score alone.

4. Conclusion

The use of scorecards in credit evaluation has become widespread in the financial industry, as they offer a systematic and reliable way to assess the creditworthiness of applicants. This study examines the predictive power of credit scores computed from a scorecard model on whether to default. In the study, a scorecard model was constructed using logistic regression in SPSS Modeler, allowing the calculation of each client's credit score. The credit score was then subjected to further analysis using LPM, logit models, and probit models in STATA. The analysis conducted through LPM, logit, and probit models consistently demonstrated that credit score has a statistically significant impact on predicting defaults. The significant relationship observed underscores the value of credit scoring models in aiding financial institutions in making informed decisions related to risk management.

However, the study used education both as an input variable in the logistic regression model to compute the credit score and as a control variable when testing the predictive power of the score on default probability. The issue of multicollinearity might occur, making it difficult to isolate the specific effect of education on default probability. Future studies should place significant importance

DOI: 10.54254/2754-1169/42/20232097

on the selection of control variables, ensuring that they are carefully chosen to avoid high correlation with the variables utilized in the calculation of credit scores.

References

[1] *Gjini, V., & Koprencka, L. (2018). Relationship Between Economic Factors and Non- Performing Loans- the Case of Albania. European Journal of Economics and Business Studies, 10(1), 253.*

[2] *Louzada, F., Ara, A., & Fernandes, G. B. (2016). Classification methods applied to credit scoring: Systematic review and overall comparison. Surveys in Operations Research and Management Science, 21(2), 117–134.*

[3] *Ripley, B. D. (1996). Pattern Recognition and Neural Networks. Cambridge University Press.*

[4] *Feng, L., Yao, Y., & Jin, B. (2010). Research on Credit Scoring Model with SVM for Network Management. ResearchGate. https://www.researchgate.net/publication/268266895_Research_on_Credit_Scoring_Model_with_SVM_for_Network_Management*

[5] *Berkson, J. (1944). Application of the Logistic Function to Bio-Assay. Journal of the American Statistical Association, 39(227), 357–365.*

[6] *Dik, V. V. (2014). Decision support methods in balanced scorecard. Naukovyi Visnyk Natsionalnoho Hirnychoho Universytetu, 4, 120–126.*

[7] *Banasik, J., Crook, J., & Thomas, L. (2003). Sample Selection Bias in Credit Scoring Models. The Journal of the Operational Research Society, 54(8), 822–832.*

[8] *Traczynski, J. (2017). Firm Default Prediction: A Bayesian Model-Averaging Approach. Journal of Financial and Quantitative Analysis, 52(3), 1211–1245.*

[9] *Bücker, M., Szepannek, G., Gosiewska, A., & Biecek, P. (2021). Transparency, auditability, and explainability of machine learning models in credit scoring. Journal of the Operational Research Society, 73(1), 70–90.*

[10] *Siddiqi, N. (2005). Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring. http://ci.nii.ac.jp/ncid/BA75777283*

[11] *Setting Options for the Binning Node. (n.d.). IBM SPSS modeler help. Retrieved June 23. 2023 from http://127.0.0.1:50727/help/index.jsp?topic=/com.ibm.spss.modeler.help/clementine/binning_settingstab.htm.*

[12] *Eeckhoudt, L., & Godfroid, P. (2000). Risk Aversion and the Value of Information. Journal of Economic Education, 31(4), 382–388. https://doi.org/10.1080/00220480009596456*

[13] *Dastile, X., Celik, T., & Potsane, M. M. (2020). Statistical and machine learning models in credit scoring: A systematic literature survey. Applied Soft Computing, 91, 106263.*

[14] *Zeng, G. (2013). Metric Divergence Measures and Information Value in Credit Scoring. Journal of Mathematics, 2013, 1–10.*

[15] *Mize, T. D., Doan, L., & Long, J. S. (2019). A General Framework for Comparing Predictions and Marginal Effects across Models. Sociological Methodology, 49(1), 152–189.*

[16] *Nguyen Viet, C. (2007). The average treatment effect and average partial effect in nonlinear models. Mpra.ub.uni-Muenchen.de. https://mpra.ub.uni-muenchen.de/44483/*