Proceedings of the 2023 International Conference on Management Research and Economic Development
DOI: 10.54254/2754-1169/21/20230274

Several Possible Causes of Credit Card Default

Mucen Ni[1,a,*]

[1]*Financial Management, Shanghai University of International Business and Economics, Shanghai, 201620, China*
a.1937119892@qq.com
**corresponding author*

Abstract: Credit card default has always existed. In this case, understanding what causes credit card default has become a matter of concern for banks and the whole society. It is necessary to understand the reasons and reduce the risk of default. This paper explores the possible causes of credit card default by studying individual credit card limit, gender, education, marriage, repayment time, bill statement and consumption amount, and combining the comparison, regression, and factor analysis in SPSS. After analysis and research, the first four possible reasons are bill statement, repayment time, consumption amount and age. In this paper, the analysis of possible causes of credit card default can help banks better understand risks, improve risk control strategies and marketing strategies.

Keywords: default, risk, analysis, bank

1. Introduction

Credit card default is a significant issue in the financial industry that can lead to negative consequences for both lenders and borrowers. The rising number of credit card defaults in recent years has drawn the attention of researchers and policymakers. A database of credit card default cases can be valuable for analyzing and predicting credit risk, detecting fraudulent activities, and developing effective strategies for reducing default rates. In that case, I choose this topic as my research direction. In this paper, I aim to analyze a credit card default database to understand the factors that contribute to credit card default. Our analysis will focus on identifying patterns and trends in the data and identifying which factors are most strongly associated with credit card default. I use the existing database for analysis in this project. Some conclusions are drawn by using SPSS comparative analysis, regression analysis and factor analysis. These conclusions will help banks and financial institutions to better understand the leading factors of credit card default in the future, and facilitate them to make some judgments after understanding customer information more effectively. There are still several questions that need to be addressed regarding credit card default. For example, what are the most common reasons for credit card default, and are these reasons consistent across different demographics? Are there any significant differences in default rates between different types of credit cards? How can lenders develop effective strategies to prevent credit card default and mitigate its negative consequences? In this work, I make several contributions to the existing literature on credit card default. I will also analyze the factors that contribute to credit card default and help banks and financial institutions to reduce the default risks.

DOI: 10.54254/2754-1169/21/20230274

2. Literature Review

Since the emergence of credit card, it has become one of the most important consumer goods in modern society. However, the widespread use of credit cards has also led to the problem of credit card default. Researchers have been exploring what factors lead to credit card default and how to prevent credit card default. For example, some studies have explored the impact of personal factors, credit card use behavior, repayment ability and debt burden on credit card default [1]. This database contains the historical data of credit card usage and default from Taiwan's bank customers, which can provide valuable insights on the issue of credit card default.

Past studies have shown that credit card default is related to many factors. Some of these factors include gender, age, marital status, education level, income level, credit limit and repayment history. So what factors mainly affect the credit card default of individual users? In this paper, I use data analysis to have an in-depth understanding of it. Different from traditional cognition, credit card default is not only related to personal economic status and credit history, but also may be affected by cultural, educational, gender and psychological factors.

3. Materials and Methods

3.1. Comparative Analysis

In this paper, I used SPSS to process the data in the database. Respectively, I used comparative analysis, regression analysis, factor analysis and so on. The first is the use of crosstabs in descriptive statistics. Fill in the columns of the form with default or not, then fill in the rows of the form with sex, education, marriage, LIMIT_BAL (amount of given credit on NT dollars) and age. I got the following tables with their respective chi-square tests. Obviously, the significance of Chi-square test is less than 0.05, indicating the validity of the experimental results. Comparative analysis uses the exploration function of descriptive statistics. Put all variables except ID and factor in the dependent variable list, and credit card default in the factor list. I made a comparative analysis in the description section. I put SEX, EDUCATION, MARRIAGE, and AGE in the cross table respectively and whether there is a credit card default. They are rows and columns in the table.

3.2. Regression Analysis

In this paper, because the database marks credit card default as 0 and 1, 0 means no default, and 1 means default. So when I face such dataset, I adopt binary logistic regression. Obviously, the dependent variable is whether the individual has a credit card default, and all the variables except ID are covariates. Then, analyze the displayed data to determine whether to reject or retain the original assumptions. The main method is to see whether the significance of the factor is less than 0.05.

3.3. Factor Analysis

The last analysis in this paper is the factor analysis. The factor analysis is crucial because it helps me accomplish the goal of this paper. I put all the other terms excluding the ID item into variables. I checked the univariate description as well as the KMO and Bartlett's sphericity test in the description, which is very important for the experimental results to be internationally recognized. In the paper, if the KMO value is greater than 0.5, of course, it is better to be greater than 0.6, it can be used. In the extraction, I chose the gravel plot to help me better select the main explanatory factors in the subsequent screening factors, and I also chose the eigenvalues greater than 1. In the rotation, I chose the maximum variance method, which is what helps me to better filter out the eligible factors. In the

DOI: 10.54254/2754-1169/21/20230274

options, I chose to sort by size as well as disable the display of small coefficients, which can be better seen and facilitate me to pick and choose factors.

4. Results

In comparative analysis, I got the following tables, including Tables 1 to 4. About the digital expression in EDUCATION, EDUCATION: (1=graduate school, 2=university, 3=high school, 4=others, 5=unknown, 6=unknown). MARRIAGE: Marital status (1=married, 2=single, 3=others)

Table 1: Data on whether men and women default

	No default	Default	Total
MALE	9015	2873	11888
FEMALE	14349	3763	18112
TOTAL	23364	6636	30000

Table 2: Data on default or not under different education levels

	No default	Default	Total
EDU 0	14	0	14
1	8549	2036	10585
2	10700	3330	14030
3	3680	1237	4917
4	116	7	123
5	262	18	280
6	43	8	51
TOTAL	23364	6636	30000

Table 3: Default data under different marital status

	No default	Default	Total
MARRIAGE 0	49	5	54
1	10453	3206	13659
2	12623	3341	15964
3	239	84	323
Total	23364	6636	30000

Table 4: Default data under different age

	No default	Default	Total
21-30	8542	2471	11013
31-40	8524	2189	10713
41-50	4606	1399	6005
51-60	1493	504	1997
61-70	189	68	257
71-79	10	5	15
Total	23364	6636	30000

DOI: 10.54254/2754-1169/21/20230274

When testing this database, I found that it did not fit a normal distribution because the histogram of the factors did not fit a bell curve. So I analyzed it using a non-parametric test. As shown in Figure 1 and Figure 2, the histograms of LIMIT_BAL and AGE in the database do not obey normal distribution. Because their histogram shapes do not fit the bell curve, as do the other factors, a nonparametric test is my method of choice.

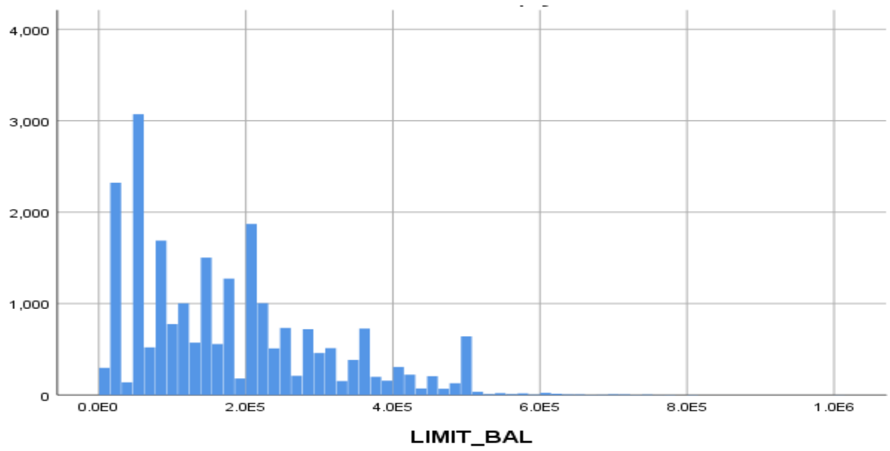

Figure 1: Histogram of the LIMIT_BAL factor

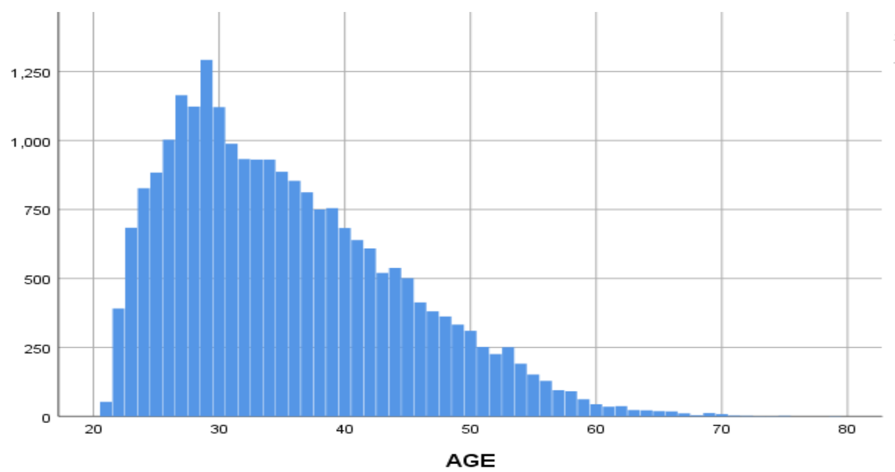

Figure 2: Histogram of the AGE factor

In the results of the binary regression, when the cut-off value is taken as 0.22, the correct percentages of default and non-default are 69.2 and 64.9, which are consistent with the data. As shown in Table 5.

Table 5: Cut-off values table

	No default	Default	Correct percentage
0	16179	7185	69.2
1	2332	4304	64.9
Total			68.3

After completing the binary regression, the boundary value is 0.22. It can be seen that gender, education level, marital status, user's loan ceiling, age, repayment status in July, August and September 2005, consumption amount in September, and bill flow in June, August and September are significantly less than 0.05, which proves that there is a correlation. However, the significance of the remaining factors was greater than 0.05, proving that the original hypothesis could not be rejected, for reasons I will explain in the discussion.

The last part is the factor analysis of the database. First, I obtained a KMO value of 0.804, which is greater than 0.6, so it proves the reliability and persuasiveness of the results. After that, in the total variance explanation, I set the eigenvalue greater than 1, so it is also convincing that the first four factors were filtered out among the eligible cases, as shown in Table 6, and could have a 65.560 percent explanatory strength for all 20 factors.

Table 6: Total variance explanation table

	Total	Initial eigenvalue variance percentage	Cumulative percentage	Total	Extracted load squared and percent variance	Cumulative percentage	Total	Rotation load squared and percent variance	Cumulative percentage
1	6.538	32.692	32.692	6.538	32.692	32.692	5.372	26.859	26.859
2	4.062	20.311	53.002	4.062	20.311	53.002	4.471	22.355	49.214
3	1.509	7.544	60.546	1.509	7.544	60.546	2.238	11.190	60.404
4	1.003	5.014	65.560	1.003	5.014	65.560	1.031	5.157	65.560

Through the rotated component matrix, we can conclude that these four factors are summarized as: BILL (bill amount), PAY (payment on time), and PAY_ AMT (monthly payment), and age. Also, the gravel plot helps me to filter the main and explanatory strength factors, referring to picture 3, and the gravel plot is more intuitive. The horizontal axis is the number of factors, while the vertical axis is the eigenvalues, so looking at the vertical axis allows me to filter out the first few factors with eigenvalues greater than 1 for subsequent generalization and analysis.

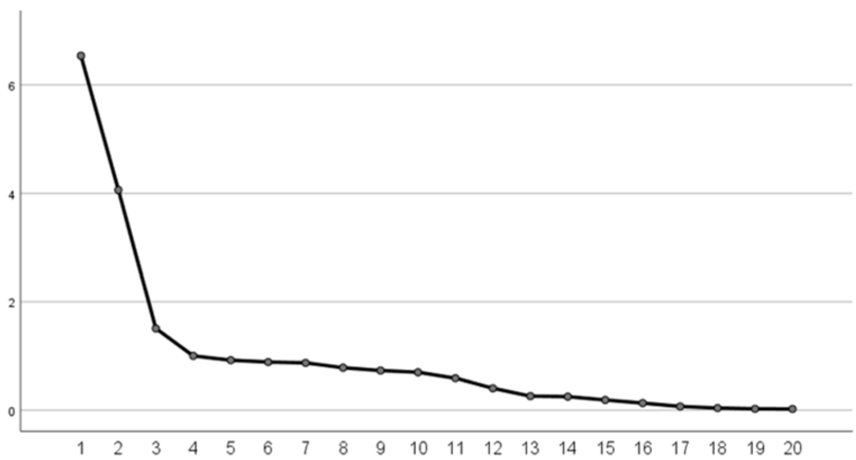

Figure 3: Gravel plot.

DOI: 10.54254/2754-1169/21/20230274

5. Discussion

First, a few tabular data from the comparative analysis, the first is gender. Through the analysis of the cross table, the first is gender. Among the genders who did not default on the next credit card payment, 61.4% were women and 38.6% were men. Among those who breach the contract, women account for 56.7% and men account for 43.3%. Both are dominated by women. Women make up the majority of those who incur credit card defaults It is because women's consumption ability is stronger than men's, so their consumption is often on the high side, so once they use credit cards, they may not be able to control their consumption, leading to the consequences of not paying back money [2]. Also, women are the majority of the non-defaulters. This is because studies have shown that women tend to do better in risk control and financial stability. Women are more inclined to budget and save, and tend to avoid high interest rate debt and high levels of consumption [3]. These are the reasons why they are in the majority of the non-default population. Next is education. Graduates, college students and senior high school students who have received education do not default in the next credit card repayment, accounting for 98.1%. But they also accounted for 99.5% of the default. People with higher education have the ability to distinguish right from wrong with cognition, so they know the impact of repayment on personal credit. At the same time, they also have more income, easier access to consumer culture, and more confidence in themselves [4]. This makes them fall into the credit card default trap. In reality, however, they may make similar mistakes to those of the less educated, such as being overly optimistic about their financial situation. Marriage is also a very important reason. Single people make up the majority of the defaulting population at 50.4%. Being single means that income comes from individuals and their spending power may be less than that of married people. But it also means that they are not bound by marriage and will spend without calculation and with themselves in mind, which often leads them to overspend or overspend, and most of them are defaulters. Married people, on the other hand, are likely to be more responsible and stable, and they are usually more focused on family and financial stability. Often have a better credit history and easier access to credit than singles, which may help reduce their credit card default rates [5]. The last one is age. Breach of contract is concentrated in people aged 22-50 years old, with a small proportion of people over 50 years old. People aged 22-50 are usually at the stage of family establishment, career development and children's education. They are under great economic pressure and spend more, so they are more likely to fall into financial difficulties. Their consumption concept may also be affected by advertising and social culture, and they are more inclined to meet consumption needs through credit cards. In contrast, people over 50 are generally in the stage of career stability and retirement preparation, and their economic situation is relatively stable. At the same time, they have more experience and knowledge about the use and management of credit cards, so they are easier to control their financial situation and avoid falling into the trap of bad debts. In addition, people over the age of 50 have generally established a good credit history, with a relatively high credit rating, and can obtain a higher credit card. Therefore, although people over the age of 50 may also have credit card defaults, the relative proportion is low [6].

The conclusions drawn after the non-normal distribution are as follows. LIMIT_ BAL refers to credit card limit. People who do not default have higher credit card limits. This is because only with good credit, which may be the history of non-default, can more limits be granted. Next is PAY_ 0-6. If the value of PAY_X equals - 1, it means the person has paid on time, so a positive value means the number of months in arrears. By studying the mean value of the two groups, we can find that the mean value of the non-defaulting group is between -0.41 and -0.21, and the mean value of the defaulting group is between 0.10 and 0.67. Both groups of data are incremental. Because the repayment has been made on time for half a year in a row, the difficulty is increasing, so the average value of non-default is close to 0, while the average value of default continues to increase [7]. Among the non-defaulting population, their average bill value increased gradually from 39,000 to 52,000, and the defaulting population also increased gradually, but the average monthly amount was smaller than the non-defaulting population. The group with larger bill amount usually has higher income and better economic strength. Their financial situation is relatively stable, and they can better manage their debts and expenses, thus avoiding the occurrence of default. In the process of using credit cards, the group with larger bill amount will usually be more cautious and rational, timely repayment and avoid excessive consumption, which will help to establish a good credit history and credit rating. PAY_ AMT1-6 represents personal credit card flow. The early payment amount of the non-defaulting group is 5,240 to 6,640, while the default group is 3,150 to 3,390. It is generally smaller than the non-defaulting group. The group with larger credit card flows usually has a higher income level, their economic strength is relatively stronger, and it is easier to pay off the credit card arrears [8]. The group with a larger credit card flows usually has a good credit history, a relatively high credit rating, and can obtain a higher credit card [9]. The groups with larger credit card flow usually have higher financial literacy and can better understand and manage their credit card use.

For the results of the binary logistic regression, I think it is a padding for the factor analysis. For example, PAY_4 to PAY_6, BILL_AMT2 to BILL_AMT6, PAY_AMT5 to PAY_AMT6, etc., they are all significant well above 0.05, so the original hypothesis cannot be rejected. But they belong to a group of variables in the group, respectively, and in the group, their respective correlations are strong. So it is common to look at a variable such as PAY_0 or BILL_AMT1 to see the effect of the factor on the original hypothesis. So their significance does not affect the final conclusion. Instead, a few significant factors or combinations can be roughly screened out in advance of the factor analysis to make some preliminary conclusions.

Finally, there is a discussion of the results of the factor analysis. First of all, the KMO value of 0.804 is convincing. For the total variance explained, I obtained four factors with eigenvalues greater than 1 and explaining 65.560% of the overall 20 variables. Second, the rotated component matrix clearly shows the percentage of each factor among the four factors filtered out. Therefore, it is concluded that these four factors are summarized as: BILL (bill amount), PAY (payment on time), and PAY_ AMT (monthly payment), and age. The biggest impact of BILL AMOUNT in credit card defaults is probably because it relates to the credit limit and debt load situation of the credit card user. A higher bill amount means that the user has a higher debt load, which could lead to a credit card default if the user is unable to make timely payments. In addition, a high bill may also reflect a user's spending habits and spending power, which could lead to a default if the user relies too heavily on credit card spending and is unable to make timely payments [10]. Payment on time can also have a significant impact on credit card defaults, which are often the result of failing to make payments on time. If a person frequently fails to make payments on time, they may be considered a high-risk customer and be subject to higher interest rates and stricter credit limits [11]. Next is the monthly payment, which to some extent reflects the consumer's ability to repay, but for those consumers who have been in arrears for a long time, their ability to repay may have reached its limit and the monthly payment amount is no longer effective in reducing the risk of credit card default. In addition, the monthly payment amount may also be affected by fluctuations in the consumer's financial situation and therefore may have a more volatile impact compared to other factors [12]. Age is usually taken into account in credit scores, which themselves are already a combination of many factors. Therefore, when other factors such as bill amount and repayment time are considered, age will have relatively less impact [13]. These are all the discussions about the conclusion. The process was not easy, but in the end, I also completed the exploration of the important reasons for credit card default. Any exploration has limitations. For example, this database has only 30000 samples, can it represent all the situations? It is obviously impossible. The database only records the half year of 2005. There is a time gap between now and now, and the cycle is also very short, which may lead to the limitation of

DOI: 10.54254/2754-1169/21/20230274

conclusions. This survey was conducted in Taiwan and may not be applicable to all regions of the world.

6. Conclusion

This paper is based on a database of credit card default. The purpose is to find out several possible factors leading to credit card default, help banks, financial institutions, etc. to effectively avoid user's credit card default, and make the most appropriate screening. Through comparative analysis, regression analysis and factor analysis in SPSS, four possible factors are obtained, namely BILL AMOUNT, PAYMENT ON TIME, MONTHLY PAYMENT and AGE. Having mastered the customer information of the above four factors can quickly determine whether a person has the possibility of credit card default. This helps the regulatory authorities better understand the credit market risks, take timely regulatory measures, and protect the stability and health of the financial market. At the same time, banks and financial institutions may also be asked to remind customers, supervise customers, avoid fines and damage to credit records caused by overdue, and help customers maintain good credit records. However, it can also be explored from gender, education, marriage and other aspects, but these factors have no absolute advantages over the first four. Of course, it also provides important research objects and data for the academic community, which is of great significance to explore the causes and solutions of credit card default and explore customer credit behavior. As the conclusion said, data-based exploration will have shortcomings. Limited data, limited time and space will limit the exploration. However, in the future exploration, efforts should be made to solve these problems, and some programs such as program design can be used to make a preliminary judgment on whether a customer will default on the credit card, which is also the direction of efforts.

References

[1] *Lin, J. Y., & Chiu, C. K. (2011). Personal finance literacy and credit card indebtedness of college students.*

[2] *Li, Y., & Yang, W. (2014). An empirical analysis of gender differences in credit card debt and repayment behavior. Journal of Consumer Affairs, 48(2), 390-406.*

[3] *Kirschenbaum, A., & Olsen, C. (2005). Gender differences in credit card debt and financial distress. Journal of Financial Counseling and Planning, 16(1), 73-85.*

[4] *Brown, M., & Graf, R. (2016). The role of financial knowledge in determining financial health. Journal of Consumer Affairs, 50(2), 269-292.*

[5] *Kim, H. S., & DeVaney, S. A. (2013). Marriage and credit use. Journal of Family and Economic Issues, 34(2), 211-222.*

[6] *Hanson, T., & Cunningham, J. (2018). Generational Differences in Financial Knowledge and Behaviors. Journal of Financial Counseling and Planning, 29(1), 92-107.*

[7] *Geng, X., Meng, F., & Xu, H. (2020). Debt Spiral, Late Payment, and Bankruptcy: Evidence from Credit Card Data. The Journal of Finance, 75(3), 1507-1547.*

[8] *Kim, J. H., & Lee, Y. J. (2016). Consumer segmentation based on credit card usage behavior: Evidence from Korea. International Journal of Consumer Studies, 40(3), 263-272.*

[9] *Lin, M., & Prabhala, N. R. (2013). Investor sophistication and credit card behavior. Management Science, 59(12), 2715-2731.*

[10] *Kim, H. Y., Lee, J. Y., & Lee, H. S. (2015). Predicting credit card default using data mining techniques: A comparison of tree-based methods and logistic regression. Expert Systems with Applications, 42(5), 2478-2489.*

[11] *Lin, W. C. (2016). An empirical study of factors affecting credit card debt payment behavior of Taiwanese consumers. International Journal of Economics, Commerce and Management, 4(1), 1-12.*

[12] *S. S. Kim and J. J. DeVaney, "Predictors of credit card payments and default among college students," Journal of Family and Economic Issues, vol. 30, no. 1, pp. 3-13, 2009.*

[13] *Agrawal, A., & Azam, M. (2016). Credit risk analysis using logistic regression modeling: A study of Indian banks. International Journal of Applied Business and Economic Research, 14(6), 4173-4193.*