Correlation analysis of telecom bank card fraud

*Yang Zhao, Ziqing Liu, Guangshuo Liu**

Qinghai Normal University, Qinghai, 810000, China

Received: 3 Jun 2025
Revised: 4 Jun 2025
Accepted: 5 Jun 2025
Published: 6 Jun 2025

Abstract*:* At present, telecommunication fraud crimes continue to occur frequently, causing huge economic losses. This study aims to explore the correlation between telecommunication fraud and given indicators, and determine whether (whether the bank card transfer transaction occurs in the same bank) and (whether the transfer transaction is online) are significantly correlated with telecommunication card fraud. Eight indicators were tested, and the results showed that all indicators did not meet the normal distribution. Then, using Spearman correlation analysis, it was determined that (the distance from the last transfer transaction) and (the ratio of the current transfer transaction amount to the last transfer transaction amount) had a low correlation with whether there was telecommunication fraud. Finally, the chi square test was used to determine the indicators that were significantly correlated with telecommunication fraud, and it was concluded that there was a significant correlation between (online transfer transactions) and whether there was telecommunication fraud.

Keywords: Telecom Fraud; K-S Test; Normal Distribution; Spearman; Chi-Square Test

1 INTRODUCTION

With the in-depth development of the information age, telecommunications network fraud has shown new trends such as diversified means, high frequency of crimes, and serious harm, becoming a prominent problem affecting social stability and public property safety. In the practice of combating telecommunications fraud, how to use data analysis technology to identify potential fraudulent behaviors has become an important issue that public security organs and financial regulatory departments need to solve urgently [1]. In order to achieve scientific identification and effective early warning of telecommunications fraud, it is of practical significance to establish a set of analysis models based on statistical and machine learning methods.

In this context, based on statistical sample data, this paper attempts to construct an analysis framework for exploring the relationship between user bank transfer behavior and telecommunications fraud. By deeply exploring multiple characteristic indicators of transfer behavior, such as transfer distance, amount ratio, whether it is an online transfer, etc., it is expected to reveal the possible inherent connection between it and telecommunications fraud [2-5]. At the same time, in order to ensure the scientificity and reliability of the research conclusions, the model construction process strictly follows the statistical test principles, clarifies the data distribution type, selects appropriate correlation test methods, and combines visualization methods to enhance the persuasiveness of the analysis [6].

The research in this paper not only provides theoretical support for identifying telecommunications fraud, but also provides a technical basis for building an intelligent anti-fraud system. By conducting a quantitative analysis of the correlation between the characteristics of transfer behavior and the probability of fraud, this article attempts to explore potential risk signals, provide decision-making references for the field of financial anti-fraud,

https://doi.org/10.71451/ISTAER2523

and promote the construction of a data-driven public security and financial collaborative governance system.

2 MODEL ASSUMPTION AND BUILDING

In this study, we built an analytical framework based on the following assumptions:

(1) It is assumed that the data in all statistical cases are true and valid.

(2) It is assumed that the method of transfer transactions in all statistical cases is certain.

To study the relationship between the first seven indicators and the occurrence of telecom fraud, statistical models can be used. Before modeling, the data should first be tested for normality distribution [7]. We use the K-S test to test the normality of the raw data, and the results are shown in Table 1, all the P values in the table are much less than 0.05, and we need to reject the null hypothesis that the raw data given in the question do not conform to the normal distribution.

Table 1: K-S test results for all indicators.

Index	K-S value	P-value	Index	K-S value	P-value
Distance1	0.3421	0.0000	Card	0.4183	0.0000
Distance2	0.4227	0.0000	Pin	0.5304	0.0000
Ratio	0.2603	0.0000	Online	0.4187	0.0000
Repeat	0.5246	0.0000	Fraud	0.5341	0.0000

Furthermore, we use the Q – Q graph to test whether the data of the first seven indicators in Shaanxi Province and whether there is telecom fraud satisfies the normal distribution Q – Q plot, as shown in the following figure 1:

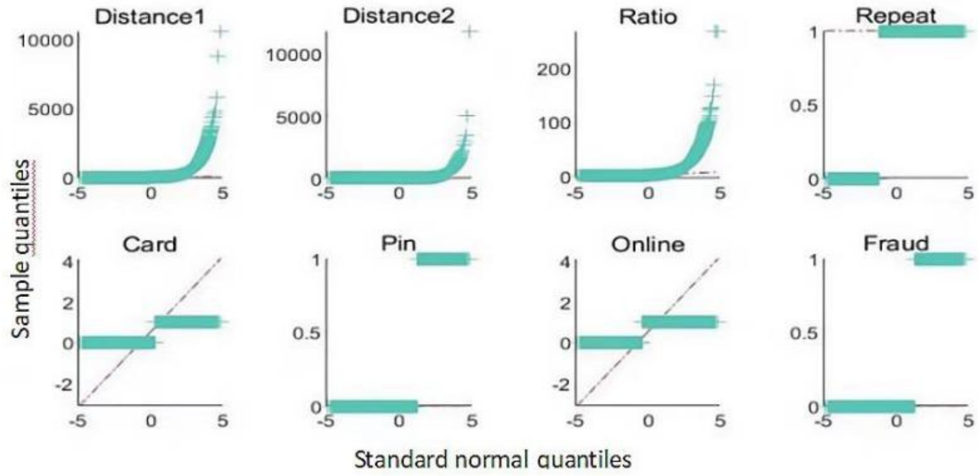
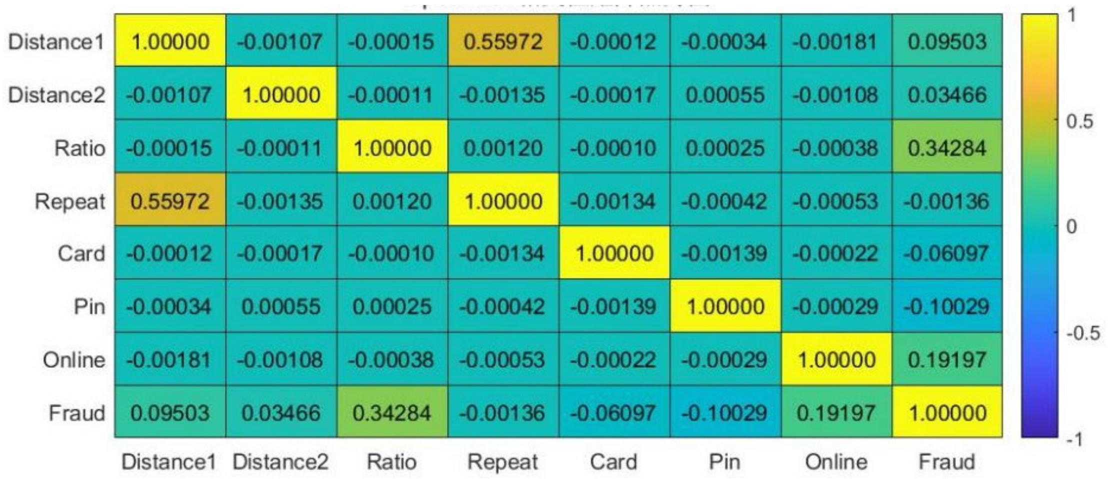


Fig. 1: Q-Q diagrams.

As can be seen from the above figure, the points on the Q – Q chart of all indicators are not approximate to a straight line, so it is considered that the data of the eight indicators do not obey the normal distribution, so we use Spearman correlation analysis [8-10]. The formula for the Spearman correlation coefficient is as follows:

$$\rho = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_i (x_i - \bar{x}) \sum_i (y_i - \bar{y})^2}} \tag{1}$$

Where ρ represents the correlation coefficient, and the closer to 1 is the more linear the correlation between the two samples. At the same time, in order to visually represent the magnitude of the degree of correlation, we draw the Spearman correlation coefficient heat map as follows:



	Distance1	Distance2	Ratio	Repeat	Card	Pin	Online	Fraud
Distance1	1.00000	-0.00107	-0.00015	0.55972	-0.00012	-0.00034	-0.00181	0.09503
Distance2	-0.00107	1.00000	-0.00011	-0.00135	-0.00017	0.00055	-0.00108	0.03466
Ratio	-0.00015	-0.00011	1.00000	0.00120	-0.00010	0.00025	-0.00038	0.34284
Repeat	0.55972	-0.00135	0.00120	1.00000	-0.00134	-0.00042	-0.00053	-0.00136
Card	-0.00012	-0.00017	-0.00010	-0.00134	1.00000	-0.00139	-0.00022	-0.06097
Pin	-0.00034	0.00055	0.00025	-0.00042	-0.00139	1.00000	-0.00029	-0.10029
Online	-0.00181	-0.00108	-0.00038	-0.00053	-0.00022	-0.00029	1.00000	0.19197
Fraud	0.09503	0.03466	0.34284	-0.00136	-0.06097	-0.10029	0.19197	1.00000

Fig. 2: Correlation analysis heat map.

We then see the correlation between the first seven metrics and the "Fraud" indicator as follows:

Table 2: Table of correlation coefficients with the "Fraud" indicator.

Index	Correlation coefficient with "Fraud"
Distance1	0.09503
Distance2	0.03466
Ratio	0.34284
Repeat	-0.00136
Card	-0.06097
Pin	-0.10029
Online	0.19197

From the above table, it can be seen that the indicators of "distance from the last bank card transfer transaction" and "ratio of the amount of the bank card transfer transaction to the amount of the last bank card transfer transaction" are low correlated with whether there is a telecommunication fraud

The chi-square test is a common statistical method used to test whether there is a correlation between two categorical variables. The P-values of "whether the bank card transfer transaction occurred in the same bank", "whether it is an online bank card transfer transaction" and whether it is a telecom fraud are obtained through the chi-side test are shown in the following table:

Table 3: Chi-square test.

Index	"Whether it is an online bank card transfer transaction" and whether it is a telecom fraud	"Whether the bank card transfer transaction occurred in the same bank" and whether it is a telecom fraud
P-value	0.000	0.176

As can be seen from the above table, the value of "whether it is an online bank card transfer transaction" and whether it is a telecom fraud is less than 0.05, so there is a significant correlation between "whether it is an online bank card transfer transaction" and whether it is a telecom fraud.

3 CONCLUSION

This study conducted a systematic data analysis and modeling exploration around the correlation between telecommunications fraud and bank transfer behavior characteristics. Through Spearman rank correlation analysis, we evaluated the correlation between multiple key behavioral indicators and whether telecommunications fraud occurred. The analysis results showed that the two indicators, "the distance from the previous bank card transfer transaction" and "the ratio between the current transfer amount and the previous transfer amount", had only a low correlation with the occurrence of telecommunications fraud, indicating that such time and amount ratio characteristics have limited ability to identify fraud in the sample.

At the same time, the remaining five indicators, such as "whether the transfer is repeated", "whether it is an online transfer", "whether the same bank card is used", etc., did not show a significant linear relationship in the correlation test with telecommunications fraud, and showed a weak correlation or even no correlation overall. This result shows that it is difficult to effectively identify the occurrence of telecommunications fraud with a single behavioral variable, highlighting the concealment and complexity of fraud in the data dimension.

To further verify the categorical correlation between variables, we used the chi-square test method to analyze the relationship between two indicators with categorical attributes, "whether it is an online bank card transfer transaction" and "whether the transfer occurs within the same bank", and whether it is telecommunications fraud. The results show that there is a significant statistical correlation between the indicator "whether it is an online transfer" and the occurrence of telecommunications fraud ($p = 0.000$), indicating that online transactions may become a preferred channel for fraudsters to some extent, and have a high risk warning value. However, the relationship between "whether the transfer occurs within the same bank" and fraud did not reach statistical significance ($p = 0.176$), indicating that this variable has limited effect in identifying fraudulent behavior, or the logic behind it has not been fully revealed in the data.

In summary, although some variables have weak or significant correlations with telecommunications fraud, overall, single-variable features have certain limitations in fraud identification, suggesting that we should further integrate multi-source information, build more complex combination variables, or adopt advanced machine learning models in the future to improve the accuracy and practicality of predictions. At the same time, the relevant findings also provide data support and strategic direction for financial supervision and risk prevention and control, especially in strengthening the monitoring of online transfer behaviors, which has positive practical guiding significance.

REFERENCES

[1] Wang, M., Xie, J., Chen, Z., Liu, M., Zhang, S., & Li, J. (2024). Enhancing Telecom Fr aud Prediction Accuracy Using a Combined CNN-LSTM Model with Bahdanau Attention Mechanism. Journal of Globe Scientific Reports, 6(2), 56-73.

[2] Wahid, A., Msahli, M., Bifet, A., & Memmi, G. (2024). NFA: A neural factorization aut oencoder based online telephony fraud detection. *Digital Communications and Networks*, *10* (1), 158-167.

[3] Liu, G. (2024). Leveraging Machine Learning for Telecom Banking Card Fraud Detection: A Comparative Analysis of Logistic Regression, Random Forest, and XGBoost Models. *Computers and Artificial Intelligence*, *1*(1), 13-27.

[4] Shao, L., Chen, Q., & He, M. (2025). Telecom Fraud Money Laundering Account Recognition Case: Multiple Machine Learning Techniques. In *AI in Banking: Practical Applications and Case Studies* (pp. 159-197). Singapore: Springer Nature Singapore.

[5] Hapase, D. S., & Patil, L. V. (2024). Telecommunication fraud resilient framework for efficient and accurate detection of SMS phishing using artificial intelligence techniques. *Multimedia Tools and Applications*, 1-23.

[6] Alshawi, B. (2024). Comparison of SVM kernels in Credit Card Fraud Detection using GANs. *International Journal of Advanced Computer Science & Applications*, *15*(1).

[7] Du, H., Lv, L., Wang, H., & Guo, A. (2024). A novel method for detecting credit card fraud problems. *PloS one*, *19*(3), e0294537.

[8] Li, F., & Chen, Z. (2025). Dynamic quantification anti-fraud machine learning model for real-time transaction fraud detection in banking. *Discover Computing*, *28*(1), 59.

[9] Charizanos, G., Demirhan, H., & İçen, D. (2024). An online fuzzy fraud detection framework for credit card transactions. *Expert Systems with Applications*, *252*, 124127.

[10] Sizan, M. M. H., Chouksey, A., Tannier, N. R., Al, M. A., Jobaer, J. A., Roy, A., ... & Aminul, D. (2025). Advanced Machine Learning Approaches for Credit Card Fraud Detection in the USA: A Comprehensive Analysis. *Journal of Ecohumanism*, *4*(2), 883-905.