Exploring the Influence Factors on Card Fraud

Tomás Fróis Duarte

Dissertation written under the supervision of Professor Nicolò Bertani, with the collaboration of Vânia Silva

Dissertation submitted in partial fulfilment of requirements for the MSc in Business Analytics, at the Universidade Católica Portuguesa, April 2024.

Contents

List of Figures

List of Tables

Abstract

Title: Exploring the Influence Factors on Card Fraud

Author: Tomás Fróis Duarte

More than 80% of non-cash payments in Portugal are made with cards. Due to this widespread usage of cards, even though fraud levels remain relatively low, fraud might imply non-negligible costs for both payment service users and payment service providers. Therefore, understanding the factors influencing card fraud is essential for formulating policies and measures by competent authorities.

This study contributes to the existing literature by conducting a comprehensive examination of the impact of both external and internal factors on card fraud.

In this study, the analysis of card fraud is approached from two perspectives: the issuer and the acquirer, both segmented into remote and non-remote transactions. In addition, cash withdrawals are included in the analysis. Fraud is quantified in this study through (i) the average fraud value, (ii) the share of fraud in volume and (iii) the share of fraud in value.

Overall, on an internal level, fraud characteristics such as type of fraud, channel type, card function, and the application of strong customer authentication exhibit significant influence on card fraud. Moreover, from an external standpoint, socio-economic factors including the impact of GDP, inflation, unemployment variation, education level, covid-19 and internet access are particularly significant in the linear regression. Therefore, implementing targeted policy measures that address these key determinants can enhance the efficacy of fraud prevention efforts and safeguard financial systems against fraudulent activities.

Keywords: Linear Regression, Statistical Significance, Calibration Study, Payment Transactions, Card Fraud

Resumo

Título: Fatores que Influenciam a Fraude com Cartões

Autor: Tomás Fróis Duarte

Mais de 80% dos pagamentos não numerários em Portugal são com cartões. Embora os níveis de fraude permaneçam relativamente baixos, devido à utilização generalizada dos cartões, a fraude pode implicar custos não negligenciáveis para os utilizadores e para os prestadores de serviços de pagamento. Consequentemente, a compreensão dos fatores que influenciam a fraude com cartões é essencial para a formulação de políticas e medidas por autoridades competentes.

Este estudo contribui para a literatura existente efetuando uma análise do impacto dos fatores externos e internos na fraude com cartões.

Neste estudo, a análise da fraude com cartões é abordada a partir da ótica: emitente e adquirente, ambas segmentadas em transações remotas e não remotas. Inclui ainda levantamentos em numerário. A fraude é quantificada através de (i) valor médio, (ii) percentagem de fraude em valor e (iii) percentagem de fraude em volume.

De um modo geral, o tipo de fraude, o tipo de canal, a função do cartão e a autenticação forte do cliente, têm uma influência significativa na fraude com cartões. Fatores socioeconómicos, incluindo o impacto do PIB, a inflação, a variação do desemprego, o nível de educação, o covid-19 e o acesso à internet, também são significativos. Desta forma, implementar medidas políticas que abordem estes determinantes-chave pode aumentar a eficácia na prevenção da fraude e salvaguardar sistemas financeiros contra atividades fraudulentas.

Palavras-chave: Regressão Linear, Significância Estatística, Estudo de Calibração, Transações de Pagamento, Fraude em Cartões

Acknowledgments

I would like to express my sincere gratitude to my supervisor, Nicolò Bertani, for his guidance, support, and encouragement throughout the research process.

I would also like to express my heartfelt gratitude to Vânia Silva for her invaluable comments and suggestions, which have significantly enhanced the quality of this work.

Additionally, I extend my appreciation to Tereza Cavaco, Hugo Mira, and Luís Oliveira, whose collaboration and support facilitated my access to the business area under study, as well as Pedro Seco for the effort in collecting and sharing the necessary information. Their cooperation was indispensable in ensuring the execution of this research.

Finally, I extend my deepest appreciation to my family and friends whose unwavering support and encouragement have sustained me throughout this journey, I am profoundly grateful.

1.Introduction

In an increasing digitalised world, ensuring not only the efficiency but also the security of electronic payments is of utmost importance. Fraud can erode trust among payers and, on the other hand, increase costs for payment service providers (PSP). Therefore, understanding the factors' influencing fraud is essential for formulating policies and measures by competent authorities.

More than 80% of non-cash payments in Portugal processed through SICOI (the Portuguese Interbank Clearing System) in 2022 were made with cards, so the analysis focuses on this payment instrument.

According to the last report of (Banco de Portugal, 2022), in the first half of 2022, cards, from the issuer perspective, maintained their position as the payment instrument associated with the highest share of fraud in both volume (0.0242%) and value (0.0228%).

It should be noted that the occurrence of payment fraud entails financial implications that may be presented either by payment service users or by PSP, establishing a threat to the security of card transactions. Consequently, it is important to understand the factors influencing the evolution of fraud levels, enabling competent authorities, such as the Banco de Portugal, to implement measures (for instance at the level of the national strategy for retail payments) aimed at promoting payment security, in line with Article 14º of its Organic Law.

To assess potential factors influencing fraud levels in card transactions, data collected by Banco de Portugal in accordance with Instruction of the Banco de Portugal no. 19/2012 was used. This Instruction governs the data collection on payment systems and instruments, namely regarding fraud, as mandated by the transposition of Directive (EU) 2015/2366 of the European Parliament and of the Council dated November 25th, pertaining to payment services in the internal market, commonly known as the revised Payment Services Directive (PSD2). Furthermore, it is in line with the guidelines defined by the European Banking Authority (EBA/GL/2018/05).

This research focuses on conducting a comprehensive analysis of detailed card payment transactions fraud. For this purpose, we investigate specific characteristics of card fraud and also explore the impact of external factors.

Leveraging historical data for pattern analysis is an effective approach. Studying the fraud characteristics, patterns and correlations will contribute to a better understanding of the panorama and gain insights for future improvements in this research field.

Employing linear regression as a supervised learning model, to study these relations, improves the model's clear interpretability, allowing straightforward understanding of how each feature impacts the outcome variable.

The analysis on card fraud was conducted from two perspectives, namely the issuer and the acquirer, categorized into remote/non-remote transactions and involved three types of dependent variables: the average fraud value and the share of fraud in value and volume.

Two groups of factors were considered as explanatory variables: (i) associated with the payment transaction and fraud characteristics; and (ii) referring to the macroeconomic context. Further in the analysis the relationship between these variables is explored. Nevertheless, two research questions have been formulated.

Research questions:

1) Which payment transaction or fraud characteristics impact card fraud?
2) Does macroeconomic context play a relevant role on card fraud levels?

On a first approach, emphasis should be placed on assessing the models' coefficients to determine their alignment with the hypothesized expectations. Additionally, attention should be given to interpreting key metrics such as the Residual Standard Error, R-squared, and Adjusted R-squared. In the context of expanding the investigation, the evaluation of regression models' predictive performance is crucial for robust analysis. Metrics such as Root Mean Squared Error (RMSE), F-statistic, and Mean Absolute Error (MAE) are effective tools for measuring forecast accuracy, particularly in short-term scenarios. Finally, in order to support conclusions regarding the significant covariates, a fraud calibration study will be developed and used to compare the actual with the predicted values.

The characteristics identified as having the greatest impact on remote card fraud are covid-19, GDP, inflation, the education level, and the type of channel used. Conversely, at the non-remote level, the same independent variables as for remote card fraud were identified as relevant, with the addition of the type of fraud. Additionally, in cash withdrawals segment, covid-19 and internet access were the components that stood out as most relevant in the regression analysis.

The subsequent sections of this dissertation are structured as follows. Chapter 2 offers an extensive review of relevant investigations and literature concerning card fraud detection models and the determinants of fraudulent behaviour. In chapter 3, a detailed explanation of the dataset respective variables is provided. Section 4 delineates the methodology employed in this study, encompassing the hypothesis of the investigation, the research model used, the data analysis component as well as the modelling techniques, and the metrics employed for result evaluation. Chapter 5 presents the empirical results resulting from the regression model, extended by a complete discussion and interpretation. Finally, section 6 outlines the major limitations identified in this investigation followed by the conclusion in chapter 7.

2. Literature Review

In recent years, the number of scientific contributions to payment fraud has increased.

Typically, this field of research focuses on two main areas: (i) the development of models for detecting fraudulent transactions, or (ii) the analysis of broader factors that may impact fraud levels. Both domains are dedicated to the prevention of fraud in payments. This dissertation aims to contribute to the latter body of literature, contemplating factors associated with payment transactions and fraud characteristics and macroeconomic elements that could influence levels of payment card fraud.

Fraud and corruption are recurring issues worldwide, impacting businesses of all types. The term “Fraud” can refer to a wide range of unethical and unlawful financial reporting practices reflecting in financial misconduct or personal gains (Jonathan M. Karpoff, 2021). However, card payment fraud is defined differently. As articulated by (Sakharova, 2012), payment card fraud constitutes the unauthorized utilization of personally identifiable information acquired by a third party without explicit consent, with the intention of attaining financial benefits.

2.1 Card Fraud Detection Models

The most effective way to reduce fraud and mitigate its associated financial damage is the early detection and prevention of fraudulent events as these are the two main mechanisms to avoid frauds and losses. Fraud prevention serves as a proactive measure aimed at preventing fraudulent occurrences, whereas fraud detection systems become operational when fraudsters circumvent the preventive measures and initiate fraudulent transactions. (Sahin, Bulkan and Duman, 2013)

Forecasting adds a practical value in auditing and government matters, paving the way for enhanced practical applications and future theoretical considerations in understanding fraud mechanisms (Kondo, Miyakawa, Shiraki, Suga and Usuki, 2019). In the realm of card fraud detection (i), the first challenge when trying to compute detection models is the difficulty in obtaining real data samples from card payments (Dal Pozzolo, Caelen, Borgne, Waterschoot and Bontempi, 2014). Secondly, the advancement of new fraud detection methodologies faces considerable challenges due to restricted exchange of ideas, particularly in the domain of credit card fraud detection, owing to security and privacy apprehensions. (Sahin, Bulkan and Duman, 2013)

Detection models trained on multiple segments exhibit higher accuracy compared to single-segment models. Multi-segment models learn from overlapping training sets; in such cases, strategies of single models can outperform the sets (Dal Pozzolo, Caelen, Borgne, Waterschoot and Bontempi, 2014).

Furthermore, detection of card fraud studies usually employs algorithms incorporating data mining methodologies. As a matter of fact, fraud detection relies on regression-based models such as Linear and Logistic Regression due to its high performance in testing factors that affects the likelihood of fraud and great explanation ability (Sharma and Panigrahi, 2012). With the exponential growth of data, data mining methods are becoming more common and crucial key tools to extract meaningful information from datasets. Fraud is one area where these techniques have shown to be incredibly effective. On the investigation of (Itoo, Meenakshi and Singh, 2020) that focused on assessing the efficacy of machine learning algorithms to detect credit card fraud, utilizing the random under-sampling method (RUS), Logistic Regression demonstrated superior performance when compared to Naïve Bayes and K-Nearest Neighbour algorithms.

On the other hand, high dimensional data and machine learning algorithms have their importance and utility. Several studies in machine learning and related fields have explored detection utilizing various methodologies including Support Vector Machines (SVMs), Artificial Neural Networks (ANNs), Decision Trees and rule-based techniques (Seera, Lim, Kumar, Dhamotharan and Tan, 2021). It is not feasible to deduce the existence of one singular optimal model or method for prediction, as various studies have demonstrated accurate card fraud prediction using diverse machine learning models combinations. In (Varmedja, Karanovic, Sladojevic, Arsenovic and Anderla, 2019) study a comparison on Logistic

Regression, Random Forest, Naïve Bayes, and Multilayer Perceptron was conducted and the Random Forest algorithm was the model that better classified whether credit card transactions were fraudulent or not. However, in a similar comparison study (Raj and Portia, 2011), that provides an overview of diverse techniques employed in credit card fraud detection mechanisms and assesses the effectiveness of each methodology, a new hybrid approach was suggested combining for instances Fuzzy Darwinian systems, neural networks, Markov models and BLAH-FDS.

Despite the existence of several detection models, it is important to note that fraud is complexly linked with external factors, resulting in constant change and evolution. Therefore, new approaches keep emerging. (Sulaiman, Schetinin and Sant, 2022) suggests a review on the detection of credit card fraud approach, focusing on a new method leveraging real-time datasets for privacy-preserving training using a Federated Learning framework with Artificial Neural Networks (ANNs) to improve fraud detection capabilities.

Furthermore, an efficient fraud detection system for cards payment fraud is indispensable for any financial institution to mitigate the threat of fraudulent transactions. Continually absorbing losses imposes significant financial losses on financial institutions, leading to heightened financial instability. Fraud prevention encompasses the implementation of anti-fraud methodologies and protocols, including but not limited to card activation, card verification codes, consumer education initiatives, address verification services, and real-time authorization protocols during point-of-sale transactions. During card-present transactions, merchants have the capability to authenticate both the cardholder's identity and the legitimacy of the card itself, thereby enhancing validation of the provided payment information by the customer. However, these measures prove less effective in card-not-present transactions (Sakharova, 2012).

The vulnerability of IT platforms has made financial institutions easier fraud targets due to their potential for large scale monetary theft through the numerous authentication flaws and loopholes within deployed serviced platform security models. The weakness in authentication methods such as signatures, PINs, passwords, and Card Security Codes (CSC), continue to be exploited by malicious third parties, enabling illegitimate transactions through innovative systems attacks (Edge and R. Falcone Sampaio, 2009).

To address vulnerabilities of those flaws in channel-level authentication mechanisms, many institutions are incorporating an additional security layer known as Fraud Management. This approach aims to overcome the shortcomings exploited by fraudsters and establishes a

comprehensive fraud control framework across all service delivery channels, creating a two-layer security model: 1) Reactive Fraud Management, which relies on knowledge discovery techniques like data mining for algorithmic processing over stored transactional data (Card-Not-Present Fraud Mitigation Framework); and 2) Proactive Fraud Management that analyses incoming transactions in real-time before authorization, identifying suspicious instances prior to any financial movement (3D Secure – a messaging protocol that strengthens the authorization of online transactions using digital certificates and passwords to authenticate both customers and payment method credentials) (Hayashi, 2020).

Additionally, the communication between the service providers and users most occur in a secure manner, ensuring robust authentication and minimizing the potential for fraud – the transaction observation mechanism should be strong enough to identify any unauthorized use of a user's personal security credentials in scenarios such as loss or theft. In that way, the article 66º of PSD2 outlines that payment service requests made by the service users via payment initiation service providers must undergo authorization using Strong Customer Authentication (SCA). The European Banking Authority has developed Regulatory Technical Standards (RTS) for SCA, facilitating reliable and secure communication among parties involved in a payment service. This technique is imperative and should be consistently applied. Regardless of how frequently a payer seeks access to their payment account or initiate transactions through a remote channel that may be fraudulent or abused, the standard authentication procedure (code creation) must be followed (Kumar Paul, 2020).

The PSP is required to request SCA whenever a customer engages in certain activities. These activities include accessing a payment account online, such as through internet banking or the PSP's mobile application. Additionally, authentication is necessary for electronic payment operations, such as in-person card purchases, credit transfers initiated through internet banking, or online transactions. Furthermore, any action conducted via remote channels that may pose a risk of payment fraud or other abuses, such as setting up recurring transfers or modifying account details, also requires strong authentication. These measures aim to enhance security and protect against unauthorized access or fraudulent activities in digital transactions (Banco de Portugal, 2019).

In other words, security should be the primary consideration for all PSP when delivering online services. Hence, in order to be able to prevent fraud occurrences it is crucial to develop strategic planning, risk detection and risk avoidance. These previous topics include monitoring internet

threats, understanding customer behaviour, and implementing robust security measures. (Fernandes, 2013).

Moreover, is important to notice that SCA is only applicable in electronic transactions. According to (ECB, 2023) exist two distinct categories of card fraud: (1) fraud occurring in card-not-present scenarios, encompassing remote transactions made online or via telephone, utilizing card information acquired through deceptive tactics like phishing; and (2) card-present fraud, commonly observed in retail stores and ATMs, entailing the utilization of forged cards.

Furthermore, the study of fraud in payment transactions has become a critical concern for PSP, especially considering the rise in fraud brought on by technological expansion and global communication. According to (ECB/2020/59) card-based payment transactions are categorized into two types: those initiated electronically and those initiated non-electronically. The rise of electronic payments (E-payments) systems has been increasing exponentially and consequently electronic frauds (E-frauds). (Fernandes, 2013). Understanding these categories is crucial for assessing the risk associated with payment fraud and implementing effective countermeasures.

Moreover, the expansion of business opportunities on the internet and the rapid growth of electronic commerce have underscored the importance of efficient payment systems over open networks. The use of traditional payment systems like credit cards encounter trust and security issues due to its reputation for insecurity and untrustworthiness caused by electronic fraud events (Abrazhevich, 2001). Hence, the author argues that the inadequacy and uncertainty of the conventional payment methods for online transactions plays a vital role in turning this environment more conducive to fraud.

2.2 Drivers of fraudulent behaviour

Card payments are directly and continually affected by external factors, encompassing not only political but also economic events such as wars and pandemics. In fact, it is expected that the variables that affect general financial fraud would also affect card payment transactions fraud, although on a different level and with a dissimilar preponderance.

Regarding the analysis of factors impacting fraud, different features could be held in consideration since the determinants of fraud are diverse, encompassing both economic and non-economic factors.

In the economic domain such as Gross Domestic Product (GDP), the Consumer price Index (CPI), unemployment, operational risk, economic freedom, and poverty have been identified as

influencers of fraud. Additionally, political factors, including governance type and political stability play a role as well as socio-economic and demographic indicators such as age, gender and occupations must be held in consideration as significant contributors. Moreover, technological changes, particularly in the context of the covid-19 pandemic, have assumed a greater importance in influencing financial fraudulent activities (Ahmad, a Ciupac-Ulici and Beju, 2021). Payment transactions exhibit a strong correlation with consumption, and consequently, with gross domestic product (GDP) that incorporates a consumption component in it. This study (Zandi, Singh and Irving, 2013) underscores the positive correlation between card usage penetration and economic expansion across diverse markets, meaning that increased card usage stimulates personal consumption and GDP growth.

According to J.M. Karpoff the effects of changing technology and gains in society's wealth may be used to create comparative predictions regarding long-term variables that are likely to have an impact on fraud. The author suggests certain adjustments that will help to raise the incidence of fraud. The development of crowdfunding platforms, for example, as well as the covid-19 pandemic and ensuing economic shutdown, have disrupted relative demands and organizational capital in ways that will likely increase the incidence of fraud in the next years. Additionally, is stated that while innovations like crowdfunding platforms may elevate fraud opportunities, advancements such as blockchain technology could mitigate fraud.

On a social level, there are certain patterns within the fraudster that should be analysed. In fact, (Nicolini and Leonelli, 2021) this study explores the role of financial literacy in preventing payment card fraud, a topic often overlooked in research. Findings indicate a positive association between financial literacy and the ability to identify fraud, particularly concerning credit card usage. Nevertheless, caution is advised against overconfidence stemming solely from financial knowledge, as it may not consistently deter fraudulent activities. Moreover, financial literacy may not be the best indicator to measure the relation between fraud and education.

Despite the existing research on (ii), there is still a need for a more thorough examination of the various factors that might affect card payment fraud.

3. Data

3.1. Dataset

To conduct this analysis, information on card payment fraud collected by the Banco de Portugal under Instruction of the Banco de Portugal no. 19/2012 was used[1]. Data was accessed through Banco de Portugal Microdata research Laboratory (BPLIM) under project ID SE_P157.

The dataset provided (FR) displays monthly observations for card issuers (i.e. entities that issue cards) and acquirers (i.e. entities that are responsible for the terminals where card transactions are acquired). Additionally, the data is divided into electronic and non-electronic transactions, with detail on remote and non-remote transactions. It also includes information on the type of fraud, as well as if strong costumer authentication (SCA) was used in the transaction.

The time span covered by the dataset comprises a period of economic and financial crisis. Covid-19 is an external factor that should not be overlooked and may influence this study specifically. Thus, when studying fraud trends within the data, is essential to consider the effect on Portuguese payments during and after the pandemic.

A second dataset (TR) with data on payment transactions for the same period was also provided. Hence TR includes fraud and non-fraud data simultaneously. This dataset was used to compute two target measures based on the share of fraud over total transactions in each period and provided the necessary groundwork for inferring about fraud across total transactions.

The data presented aligns with the overview of card fraud delineated in the payment system report, issued by Banco de Portugal in 2022. According to the report and the data provided, the share of fraud in domestic transactions, which accounted for 84% of all transactions, is more substantial in cards from the issuer's point of view. Comparing the share of fraud in the first half of 2022 with the same period in the previous year, there was a decrease that should be noted. In terms of share of fraud by volume, fraud fell from 0.0261% to 0.0242%, in terms of value from 0.0333% to 0.0228% and the average value per fraudulent transaction fell from 54 to 45 euros. From the buyer's point of view, the figures also varied. The share in volume went from 0.0003% to 0.0002%, in value from 0.0014% to 0.0016% and the average value per fraudulent transaction went from 230 to 308 euros. The variation in these metrics may be based on various types of external factors, which this dissertation aims to focus on.

[1] The analyses, opinions and findings expressed in this dissertation represent the views of the authors and not necessarily those of Banco de Portugal or the Eurosystem.

Furthermore, is important to notice that card fraud amount has been increasing from 2020 onwards. Most of this amount is set in 2022 with a total of around 26 million euros, followed by 2021 (22 million euros) and 2020 (20 million euros). The amount of fraud has been increasing over the last 4 years. In 2023 it is expected that the value of fraud will not exceed the one in 2022, since the figures up to August 2023 are about twice as low as in the homologous year.

3.2. Explanatory Data Analysis

For the purpose of this study, exclusive focus was placed on fraud in payments that include cards, thereby enabling segmentation by card issuer, card acquirer, and cash withdrawals. The data provided by Banco de Portugal ranges from January 2020 to August 2023 and contains several monthly reports by reporting entity. The number of observations during this time interval is 35 101 and includes 13 variables.

Variable Name	Variable Type	Description
Year	Numeric	Year of fraud report. (2020-2023)
Month	Numeric	Month of fraud report. (1-12)
Country	Category	Location of the POS (Point of sale) terminal.
Instrument	Category	Payment instrument. (1 – Card (acquirer), 2 - Card (issuer), 3 - Cash Withdrawals)
Electronic	Binary	Non-Electronic or Electronic payment. (0: Non-Electronic, 1: Electronic)
Remote	Binary	Non-Remote or Remote payment. (0: Non- Remote, 1: Remote)
SCA	Binary	Indicates if SCA was used in the payment. (0: Non- SCA, 1: SCA)
Motive_N_SCA	Category	Reason for not using SCA according to the RTS. (Regulatory Technical Standards)
Fraud	Category	Type of fraud according to the EBA Guidelines.
Funtion_Card	Category	Type of card function used. (1 - Debit, 2 - Credit, 3 - Delayed Debit Card)
Channel	Category	Type of Channel used. (1 - ATM or other PSP terminal, 2 - Mobile payment solution, 3 - Initiated at a physical EFTPOS, 4 - Other)

QT	Numeric	Quantity of fraud reported (in units).
MT	Numeric	Amount of fraud reported (in euros).

Table 1 – Data Dictionary

The data is not homogenously distributed throughout the payment instruments and may condition the analysis, so this factor is taken as a limitation. In fact, cards from an issuer perspective comprise 92.2 % of the observations taken into consideration.

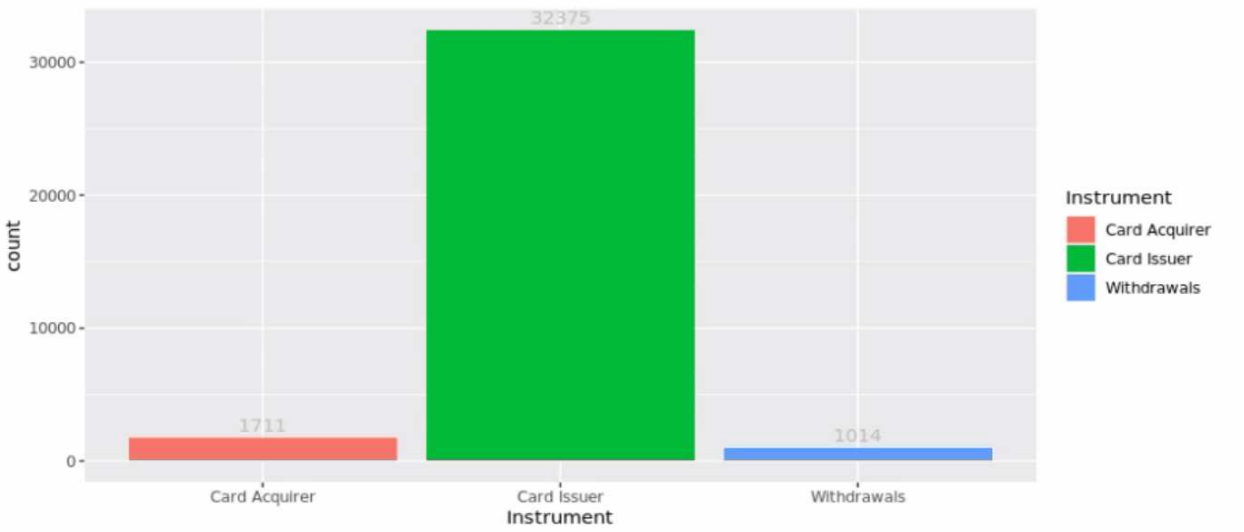

Figure 1 - The distribution of observations across the payment instruments

As evidenced in Figure 2, 3 and 4, there is a significant discrepancy, proportionally, between the cards from the issuer's perspective compared to the acquirer's perspective and cash withdrawals. These figures show the evolution of the dependent variables over time. Figure 2 demonstrates the average monthly value of card fraud per PSP by instrument over time. In this graph, it is observed that the average value of fraud is notably higher in cards (issuer), followed by cards (acquirer) and cash withdrawals.

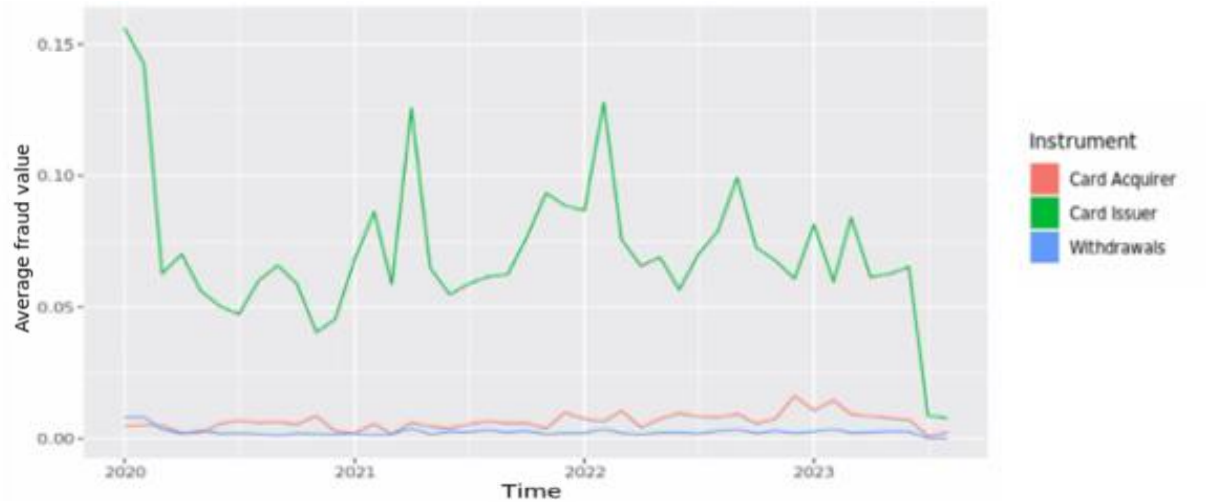

Figure 2 - Average fraud value by instrument over time

Figures 3 and 4 depict the evolution of fraud shares in both value and volume. Two peaks are noticeable at the beginning of the year, both in 2021 and 2022. This highlights potential correlations of fraud with external events or seasonality. Therefore, the peaks observed in the first semester will be subjected to temporal analysis, considering their observed pattern.

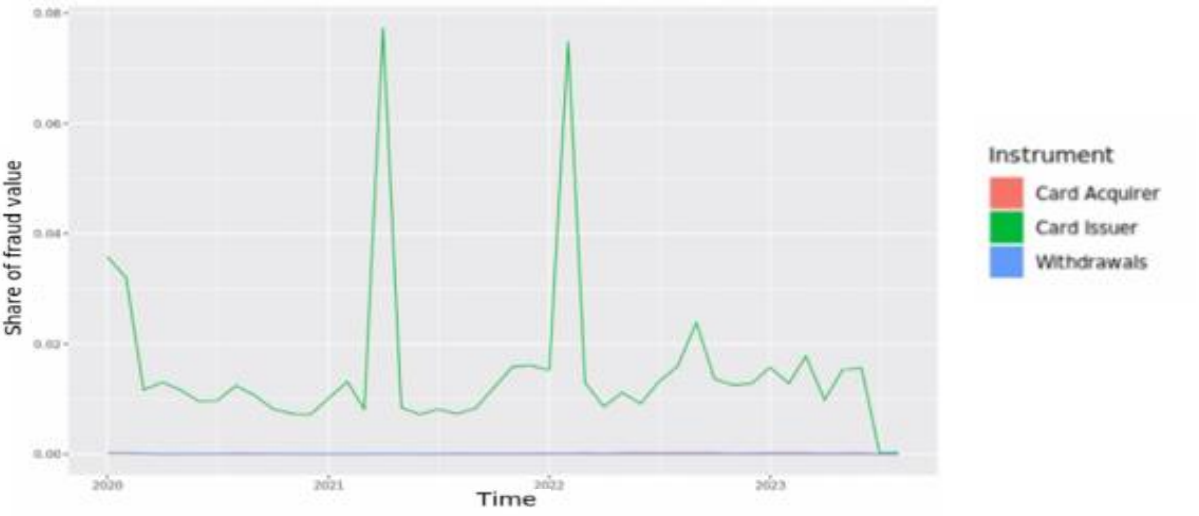

Figure 3 - Share of fraud value by instrument over time

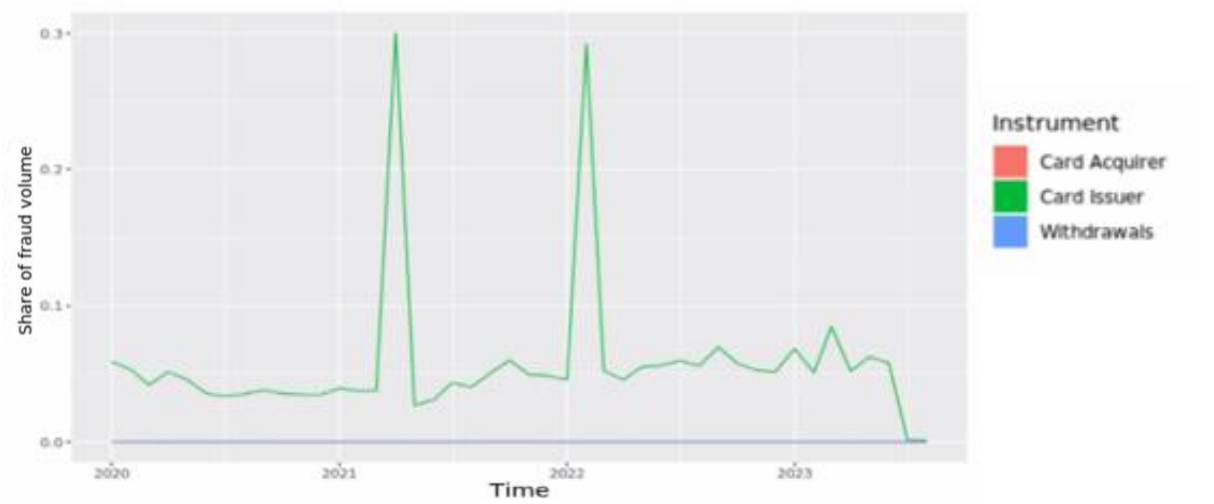

Figure 4 - Share of fraud volume by instrument over time

The visual inspection of the data reveals a pattern indicating elevated fraud values during the end-of-year and beginning-of-year periods. This trend is reasonably anticipated, as increased consumer activity during these periods typically correlates with a greater frequency of transactions, thereby presenting more opportunities for fraudulent activities to occur.

Electronic payments account for the majority of card fraudulent transactions. For instance, cards in an acquirer perspective have 88.43% of their fraud being electronically, while cards in an issuer perspective have 64.25%. Conversely, cash withdrawals are excluded from this representation as well as from the remote, SCA and type of channel used, due to their classification outside the mentioned payment realm.

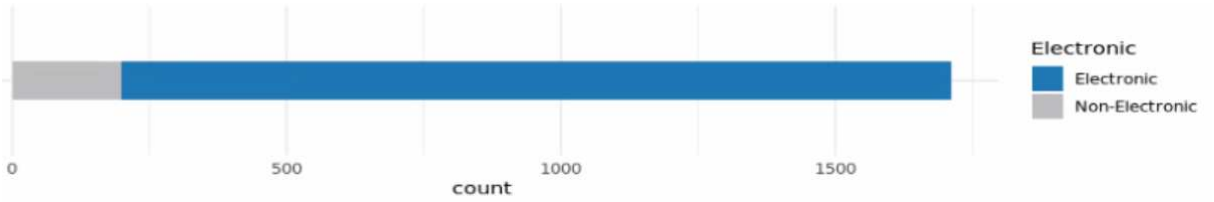

Figure 5 - Distribution of Fraudulent Electronic Payments in Cards (Acquirer)

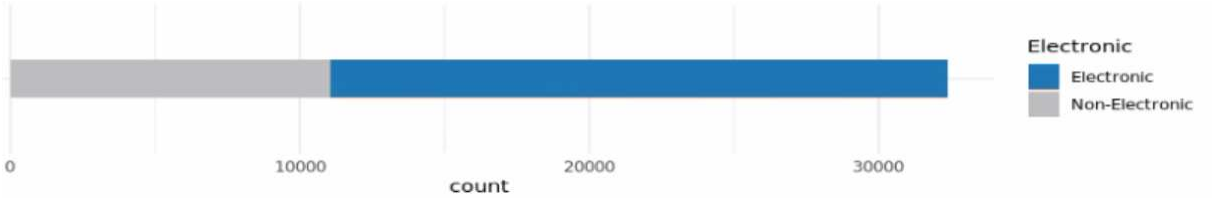

Figure 6 - Distribution of Fraudulent Electronic Payments in Cards (Issuer)

Furthermore, the fact that 58.1% of fraud on issuer cards is remote suggests that a substantial portion of fraudulent activities occurs without the physical presence of the card. On the other

hand, with 31.9% remote fraud in acquirer cards, while still notable, indicates a lower prevalence of such activities compared to issuer cards.

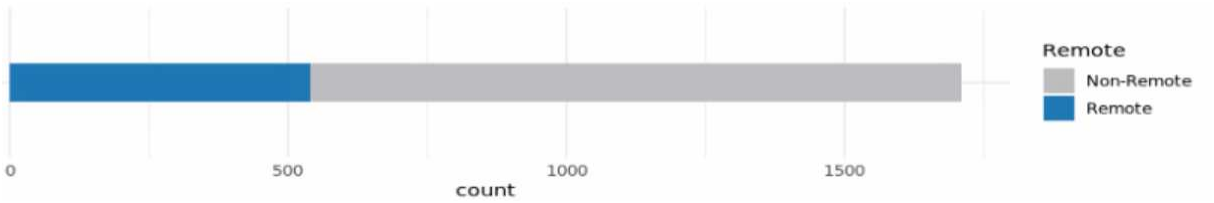

Figure 7 - Distribution of Fraudulent Remote Payments in Cards (Acquirer)

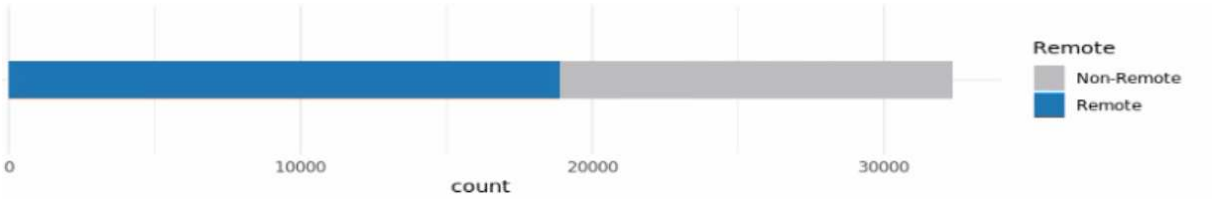

Figure 8 - Distribution of Fraudulent Remote Payments in Cards (Issuer)

An essential factor to highlight is the implementation of strong customer authentication, enabling individuals to authenticate the validity of a particular payment. In card transactions where SCA was applied, the fraud rate was significantly lower compared to transactions without this type of authentication. Therefore, it is anticipated that there exists an inverse relationship between this variable and the occurrence of fraud. Only 20.1% of fraud in acquirer cards involved strong authentication, whereas in issuer cards, the figure was 12.17%.

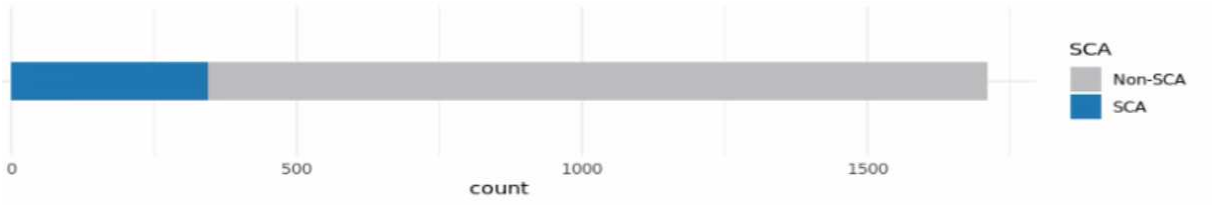

Figure 9 - Distribution of SCA Payments in Cards (Acquirer)

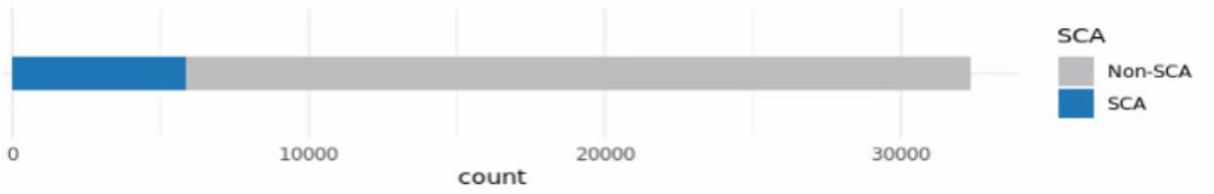

Figure 10 - Distribution of SCA Payments in Cards (Issuer)

The direction of card fraud flow is contingent upon both the payment service utilized and the initiation channel employed. Among the various initiation channels reported, the most frequently encountered is transactions initiated at a physical EFTPOS (Electronic Funds Transfer at Point of Sale) terminal. EFTPOS terminals, which electronically capture payment data, are capable of transmitting information either online, through real-time authorization requests, or offline, as defined by the European Central Bank regulation (ECB/2020/59).

Figures 11 and 12 suggest that 62.6% of card (acquirer) fraud is originated at physical EFTPOS terminals, while in the issuer card perspective 57.1%. The remaining majority involved fraud in other types of channels besides ATMs or mobile payment solutions.

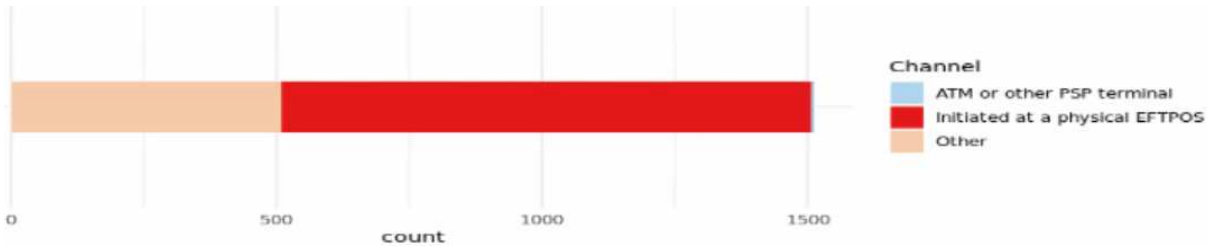

Figure 11 - Distribution of the Type of Channels in Cards (Acquirer)

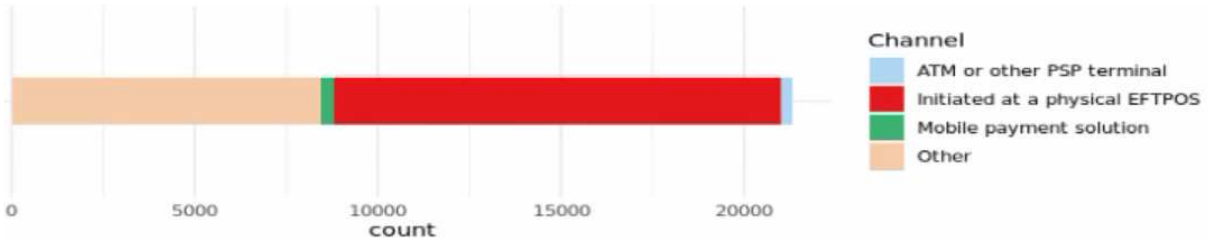

Figure 12 - Distribution of the Type of Channels in Cards (Issuer)

In addition, debit and credit cards are common types of payment cards utilized for various financial transactions, consequently, they are the most frequently employed payment methods for fraudulent activities. By the analysis of Figure 13, we can infer that 73.6% of the fraud was reported on credit cards, while in Figure 14, the scenario is reversed, with debit cards accounting for a higher value of fraud at 51.6%, followed by credit cards at 47.6%. In cash withdrawals, the distribution is homogeneous between debit at 55.38% and credit with the remaining percentage.

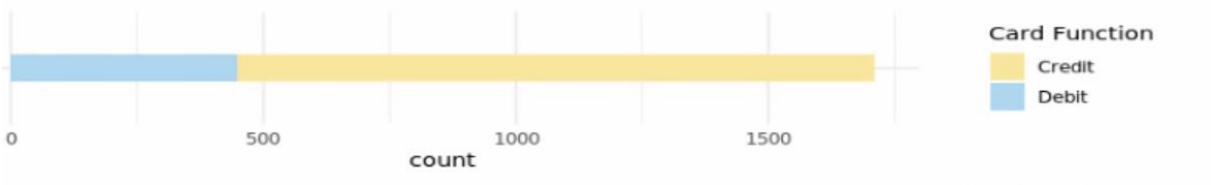

Figure 13 - Distribution of the Types of Cards Function in Cards (Acquirer)

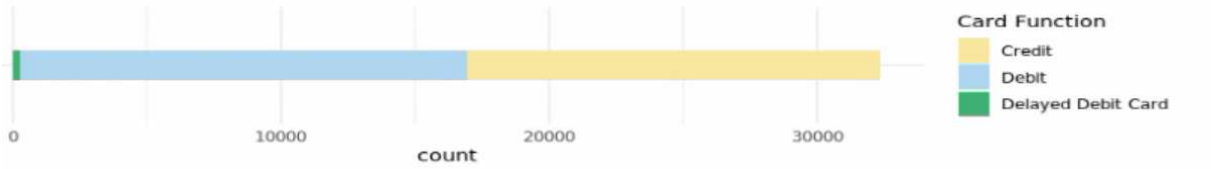

Figure 14 - Distribution of the Types of Cards Function in Cards (Issuer)

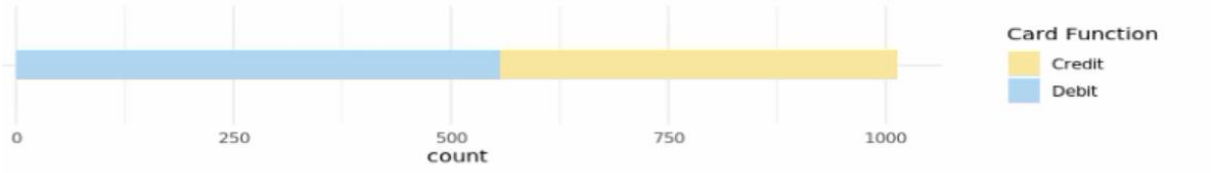

Figure 15 - Distribution of the Types of Cards Function in Cash Withdrawals

The type of fraud variable follows the "Final Report on Fraud Reporting guidelines under PSD2" document from the European Bank Authority (EBA/GL/2018/05). This document classifies three main types of fraud and groups them within the possible fraudulent sub-category. Fraud in payment transactions can be divided generally in three categories: 1) Issuance of a payment order by the fraudster , including lost or stolen cards, cards that were not received, counterfeit cards, unauthorized payment transactions, including those resulting from the loss, theft or misuse of sensitive payment data or payment instrument ; 2) Manipulation of the payor , meaning payment operation carried out when the account holder is deceived by the offender to issue a payment order or instructions to the PSP, in good faith, believing that the destination account belongs to the legitimate beneficiary; and 3) Modification of a payment order by the fraudster (EBA article 96º, nº6, from PSD2).

Type of Fraud	
Issuance of a payment order by the fraudster	Lost or stolen card
	Card not received
	Counterfeit card
	Card details theft
	Other
Modification of a payment order by the fraudster	-
Manipulation of the payer by the fraudster to issue a payment order	-

Figure 16 - Type of fraud subdivision

The most common type of fraud in card payments is the issuance of a payment order by the fraudster. This type of fraud accounts for approximately 83% of card fraud (acquirer), 88.4% in card (issuer) fraud, and 91.8% in cash withdrawals. The occurrence of "Manipulation of the payer by the fraudster to issue a payment order" typically demonstrates a lesser prevalence when contrasted with the type "Modification of a payment order by the fraudster," barring exceptions noted in cash withdrawals. This pattern is anticipated due to direct interactions with the fraudster in such in-presence interactions.

Figure 17 - Distribution of the Type of fraud in Cards (Acquirer)

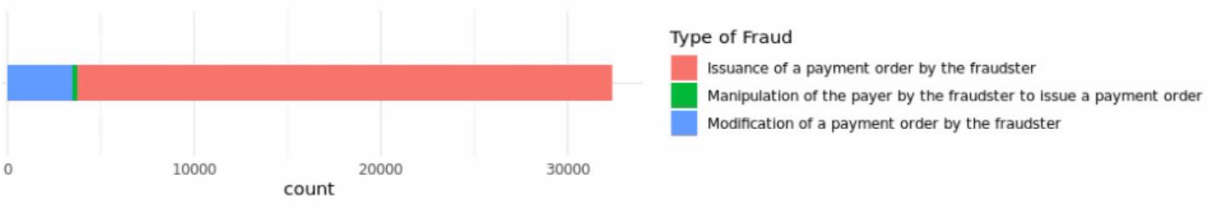

Figure 18 - Distribution of the Type of fraud in Cards (Issuer)

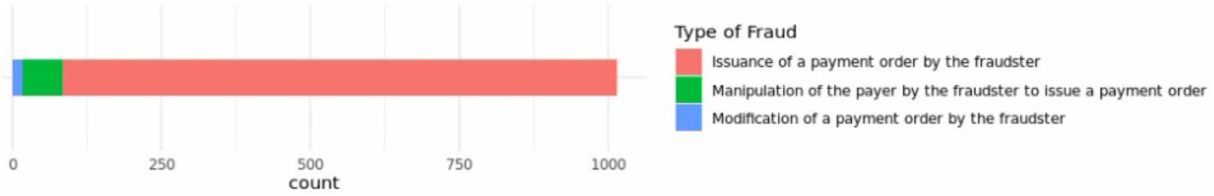

Figure 19 - Distribution of the Type of fraud in Cash Withdrawals

4. Methodology

Attempting to address the initial research questions, a four-phase methodology was employed and two null hypotheses where made:

> *H0*-1: There is no association between payment transaction characteristics and card payment transactions fraud.
>
> *H0*-2: Economic conditions do not contribute to the study of the influence factors on card fraud.

The four-phase methodology steps collectively ensured data integrity and enriched the dataset with relevant information, facilitating a comprehensive analysis of fraudulent activities in card payment transactions.

Firstly, feature engineering techniques were implemented to preprocess the information. This involved handling missing values and outliers' removal, employing one-hot encoding for categorical variables, and normalizing numerical features using re-scaling methods.

Secondly, a detailed analysis was undertaken, focusing not only on time series changing aspects but also on individual card instrument categories. This approach enabled the identification of nuanced patterns and specific trends leading to an identification of the logical features to incorporate in the models.

When computing the linear regression models theoretically considerations and external variables were incorporated based on patterns identified through the temporal analysis and also economic studies regarding the corruption influence factors. Furthermore, an exclude-variable selection was developed in order to avoid perfect-collinearity or multi-collinearity in the linear models. Finally, a calibration test was conducted by comparing the predicted results generated by the models with the actual observed outcomes.

4.1. Research Model

Considering the aim of this study and the research questions that focus on the influence of both internal and external factors on fraud, linear regression has been selected as the model for investigation.

The Linear regression model, as a supervised model, is a fundamental statistical method used to compute the relationship between a dependent variable *y* and one or more independent variables $x_1, x_2, \ldots, x_n$. As stated on the following multilinear regression model formula.

$$y = B_0 + B_1 x_1 + B_2 x_2 + \cdots + B_n x_n + \epsilon$$

The concept of univariate regression investigates the relation between a solitary dependent and independent variable, examining a one-to-one relationship, whereas multivariate regression analysis the correlation between a singular dependent variable and multiple covariates (Peja, 2023). Given the complexity of the dataset and the presence of multiple factors potentially influencing the dependent variable, multivariate regression is the preferred approach for this analysis.

In addition, linear regression coefficients $B_0, B_1, \ldots, B_n$, represent the change in the dependent variable for a one-unit increase in the independent variable, holding other variables constant. Their sign indicates the direction of the relationship. These coefficients are associated with *p-values*, serving as indicators of the statistical significance of the relationship between each independent variable and the dependent variable. The *p-value* quantifies the strength of evidence against the null hypothesis. According to (Wooldridge, 2012) small *p-values*, generally below 5%, are evidence against *H0* because, if *H0* is true, the indicator show that the data's outcome occurs with a low probability. Therefore, the smaller the *p-value*, the stronger the evidence against the null hypothesis, leading to its rejection.

The error term ϵ holds what remains unexplained by the independent variables in a regression, meaning that this term captures influences on *y*, that are not accounted by the independent variable *x*. Additionally, it is presupposed that the residuals adhere to a normal distribution, with the implication that the expected value of ϵ equals zero and exhibits a constant variance, even if partially differentiated (Peja, 2023).

Within the same topic, is important to mention the Ordinary Least Squares (OLS) methodology. OLS is a statistical method that can be used in linear regression for estimating parameters. The objective of using this model is to reduce the total squared discrepancies between the regression equation's predicted values and the dependent variable's observed values. In order to achieve valid results, the mathematicians Carl-Friedrich Gauss and Andrei Markov developed several key assumptions for the linear regressions (Hayes and Cai, 2007), such as:

i) Linearity: This assumption states that the dependent variable has a linear relation with the parameters and is generated according to the previous regression formula.
ii) Random Sampling: This assumption ensures that the sample is representative of the population and the true population is the center of the sampling distribution.
iii) No perfect collinearity: There is not always a perfect correlation between independent variables.
iv) Exogeneity: Th independent variables remain unaffected by the error term within the model, indicating the absence of correlation between covariates and the residual.
v) Homoscedasticity: This assumption refers that the variability of the residuals remains regular across all levels of the independent variables, ensuring low variance and stability in the model's predictability.

An extension of Ordinary Least Squares (OLS) regression that offers a way to handle heteroskedasticity is Weighted Least Squares (WLS) regression. According to (Hayes and Cai, 2007), while OLS weight all observations equally, WLS allows researchers to assign weights to individual cases based on the variance of the errors. In fact, using this method, standard error estimation can be improved, as well as statistical inference accuracy by considering heteroskedasticity. Additionally, Generalized Least Squares (GLS) method is also a method that, similar to WLS, account for heteroskedasticity. There difference relies on the approach, while WLS adjust the weights to individual observations, GLS models the covariance structure of the errors by specifying a covariance matrix of the errors.

4.2. Data Preparation

After analysing the data, changes that needed to be made were identified so that the data could be fed into models and their bias reduced.

In order to avoid skewing the interpretations of the data obtained, statistical detection methods, more specifically Gaussian-based methods, were used to analyse outliers in the numerical variables. Boxplot and mean–variance are the most commonly used techniques in Gaussian-based methods (Smiti, 2020). There were several changes due to the outlier's treatment, such as the decision to reject the inconsistent value of 0.00 euros in the amount reported, as it is required that all reports exhibit positivity.

Additionally, as part of the data preparation process, one-hot encoding techniques were employed. According to (Rodríguez, Bautista, Gonzàlez and Escalera, 2018) the most common method for handling multi-class classification is still one-hot encoding. This approach aimed to

address multicollinearity issues and effectively mitigate the dummy variable trap by transforming categorical variables into binary vectors.

Fraud can be perceived and measured in distinct ways, in this analysis is considered three measurement systems to answer the research questions: the average fraud value and the share of fraud in value/volume. This metrics do not directly exist in the dataset; hence, they were computed:

1) Average Fraud Value: Amount of Fraud (MT) reported by a PSP over the Quantity of Fraud (QT) reported by the same PSP.
2) The Share of Fraud Value: Amount of Fraud reported by a PSP over the Total Amount of Transactions in that same period (MT FR / MT TR).
3) The Share of Fraud Volume: Quantity of Fraud reported by a PSP over the Total Quantity of Transactions in that same period (QT FR / QT TR).

From an economic point of view, relevant events that occurred during the review period might have had an impact on payment transactions. In addition to the outcome variable, new economic and sociodemographic covariates were created to support the analysis:

1) Covid-19: People avoided in-person interactions and favoured remote purchases during the pandemic, which considerably accelerated the shift towards online transactions and digital payments (Sharma, 2020). Portugal reported the first victim of virus SARS-CoV-2 in March 2020, so a binary feature with a milestone on this date was added to the dataset. In response to the evolving pandemic in 2020, Banco de Portugal in collaboration with the Portuguese banking community, increased the maximum allowable amount for contactless payments without requiring the entry of a PIN. The limit has increased from 20 to 50 euros per transaction (Banco de Portugal, 2020). At the economic/technological level, this measure enabled higher-value contactless transactions, potentially increasing instances of card theft and other fraudulent crimes. Therefore, the variable "Covid-19" encompasses this factor within its scope.
2) Ukraine's military conflict: Cross-border transactions and international payment patterns may have been impacted by trade tensions and geopolitical uncertainties. The projections, in a short-term outlook for the Portuguese economy are negative. The year 2023 will see the expected effects with a significant slowdown in economic activity in comparison to 2022 (Banco de Portugal, 2022).

3) Δ Gross domestic product: As GDP serves as an indicator for economic activity and financial stability, it is conjectured that changes in GDP levels may correspond to changes in the average value of fraudulent activities. The variation in GDP acts as an indication of welfare and advancement. (Dimitras, Kyriakou and Iatridis, 2015).
4) Δ Inflation: This phenomenon, characterized by a sustained increase in the general price level of goods and services, directly correlates with the reduction of consumers' purchasing power, consequently leading to a decrease in payment transactions. Not all payment instruments may correlate with inflation, however, it is conceivable that electronic money transactions may significantly impact inflation dynamics. (Titalessy and Bethania, 2020)
5) Education: To measure this variable, the applied metric was the number of students enrolled in tertiary (post-secondary) education. On the socio-demographic level, it is expected that a higher level of education correlates with a lower crime rate. (Lochner and Hjalmarsson, 2012)
6) Δ Unemployment: Annual changes in the unemployment rate is an impactful sociodemographic variable to include. As stated in (Ahmad, Ciupac-Ulici, and Beju, 2021), during economic contractions, unemployment demonstrates a notable positive correlation with crime, whereas crime rates decline during periods of economic expansion.
7) Internet Access: Higher percentages of households with internet access and broadband connections may lead to an increase in online transactions, including card payments, as well as an increase in the lifestyle of the families, increasing their payments. Credit card fraud poses an escalating risk to businesses engaged in online sales of goods or services (e-commerce) as mentioned in (Delamaire, Abdou and Pointon, 2009).

Variable Name	Variable Type	Description	Source
PT_terminal	Binary	Location of the POS terminal. (0: Not in Portugal, 1: Portugal)	FR dataset
Covid	Binary	Occurred during covid. (0: No 1: Yes)	-
Ukr_war	Binary	Occurred during Ukraine's conflict. (0: No, 1: Yes)	-

Gdp_pt	Numeric	Annual Gross Domestic Product variation in Portugal.	Eurostat
Inf_pt	Numeric	Annual Inflation variation in Portugal.	Eurostat
Edu_pt	Numeric	Number of Portuguese students enrolled in tertiary education.	Eurostat
Unemp_pt	Numeric	Annual Unemployment variation in Portugal.	Eurostat
Internet_pt	Numeric	Private households with a computer internet access, and broadband internet connection (%).	Pordata

Table 2 – Data Dictionary 2 - Details of the new features created

Ultimately, the final transformation applied to the data was a normalization method. This approach proves beneficial, especially in linear regressions. The objective of normalization techniques is to map the data to a consistent range. Given that the numerical variables within this dataset exhibit varying units and scales, Min-Max scaling was implemented. The primary advantage of Min-Max normalization lies in its ability to constrain all values within a predefined range. Moreover, Min-Max exhibits fewer misclassification errors in comparison to Z-Score or Decimal Scaling methods. (Saranya and Manikandan, 2013). Nevertheless, it may pose challenges to the interpretation of the results.

4.3. Time Series Analysis

When examining time factors, different conclusions can be drawn depending on the perspective (issuer, acquirer, or cash withdrawals). There are some irregularities with the data, however, between the months of December to April the overall value of *y* increased, especially in cards (issuer). This period encompasses events characterized by significant payment transactions, such as Christmas, New Year, and sales seasons. In this part of the year payment transactions increase and as expected fraud as well. On the other hand, the following months are characterized by a decrease in consumption.

The temporal analysis was conducted separately for each card type, considering the remote and non-remote components. Regarding fraud involving remote cards from the acquirer's perspective, there is a discernible trend: a decline from 2020 until the early months of 2021, followed by an upward trajectory thereafter. Additionally, seasonal patterns are observed, with

recurring peaks at specific times each year. For instance, there is a notable rise in remote fraud at the beginning of each year, consistent across all years under analysis. Particularly noteworthy is the data from 2023, when the value doubled compared to the same homologous period. Furthermore, the monthly lags displayed in both the ACF and PACF correlograms, as detailed in the Appendix, are statistically significant for the month of January.

Fraud from the same perspective but on non-remote cards is also indicative of a trend and seasonality. The trend is increasing throughout the analysis while the seasonality is, as in the remote case, annual. The ACF and PACF graphs for this segment have significant autocorrelation at the first legs indicating that there is a strong correlation between adjacent observations in the time series. The random component represents fluctuations in the data that cannot be attributed to the seasonal or trend components. In this case, since the random component doesn't show a pattern, it suggests that the seasonal and trend components mostly capture all the underlying patterns in the data. Otherwise, if the random component shows a pattern or structure, it could suggest that the seasonal and trend components do not fully capture all the underlying patterns in the data.

In the case of issuer-remote cards, the annual seasonality is evident, with a pattern characterized by a peak in fraud at the beginning of the year, followed by a sharp decrease and consequent reduction in variance in the following months. Once again, it is the first legs of the ACF and PACF plots that are significant, and nothing can be inferred about the remaining ones. There is also no clear trend for the non-remote segment, unlike annual seasonality. The behaviour exhibits remarkable similarity between remote and non-remote fraud from the issuer's perspective.

Cash withdrawals display annual seasonality characterized by patterns with relatively high variance. The stochastic component in this payment instrument cannot be entirely attributed to seasonality and trend, indicating the presence of external factors significantly impacting withdrawals. Considering the time frame, it is anticipated that the covid-19 pandemic may have contributed to this behaviour.

After the time series analysis, we can infer that there is significant autocorrelation at lag 1 in every payment instrument (all cards acquirer/issuer perspective and cash withdrawals). Therefore, the previous OLS assumption iv) Exogeneity is being violated because it implies a correlation between the current error term and previous error terms, indicating that the error term within the model does not entirely affect the independent variables.

4.4. Modelling

In the modelling process was computed five linear regressions with different features according to their disaggregation and the external factor impacting the outcome.

The variables included in each regression are theoretically grounded based on how they are reported by PSP. In other words, depending on the type of fraud and the perspective, there are characteristics and variables that do not apply. For example, SCA only applies to electronic payments, and the same pertains to the type of Channel, Type of Fraud, and the reason for non-application of SCA. This approach ensures that the model's parameters are estimated exclusively based on available and reliable data.

Time series decomposing techniques were implemented with the objective of analysing the components of trend, seasonality, and randomness. Additionally, the autocorrelation function (ACF) and partial autocorrelation function (PACF) methodologies allowed for a better understanding of specific periods (lags), and it further elucidated the temporal dependencies between observations. ACF displays the correlation coefficients between the original time series data and its lagged values at different time lags. On the other hand, PACF isolates the correlation between a data point and its lagged values, excluding the indirect influence of other lags by controlling for the effect of intermediate lags.

This process involved incorporating, in all card segments, a temporal variable that can catch the seasonality inherent in the data, confirmed by the continuous significance in lag 1 of the ACF in the Appendix, and, additionally, fill the expectation of an extra consumer behaviour in this period of the year. Therefore, an interaction term "nov_dez:dez_jan" was added to each linear model as a seasonal variable that reflects how the fraud changes in the transition from November to December and then to January.

Overall, the linear regressions integrate insights gathered from a multifaceted analytical approach, including explanatory analysis, time series analysis, and new standardized socio-economic features that, according to the literature, may influence fraud corruption. Moreover, in order to avoid perfect collinearity, variables that indirectly bias the models (MT and QT) were removed since they are already being accounted in the dependent variable formula. Multicollinearity refers to a statistical phenomenon in regression analysis where two or more predictor variables in a model are highly correlated with each other. In fact, the incorporation

of economic variables encompassing other factors within the same regression model can introduce multicollinearity issues. Therefore, a feature selection technique grounded in correlation analysis, statistical and theoretical significance was implemented. This method final objective is to answer the research questions, by prioritizing variables that contribute significantly to the outcome of interest. Therefore, this approach enhances the precision and interpretability of the regression model, enabling a better understanding of the factors influencing fraud (Li, Cheng, Wang, Morstatter, Trevino, Tang and Liu 2017).

With the intention of verifying the accuracy of the model, a prediction was conducted. This forecast was made to evaluate how well the models could predict or estimate results. To conduct the assessment the dataset was partitioned into an 80-20 split. The training set, using 80% of each card segment observations, was utilized to fit the model while the testing set with the remaining 20% reserved to evaluate the prediction comparing the observed with the actual values of the dependent variables. Typically, this data split process avoid overfitting the data, otherwise the model could lead to less accurate predictions.

	Number of observations in the train.	Number of observations in the test.	Total number of observations.
Cards (Acquirer) - remote	432	108	540
Cards (Acquirer) - non-remote	936	235	1 171
Cards (Issuer) - remote	15 093	3 774	18 867
Cards (Issuer) - non-remote	10 806	2 702	13 508
Cash Withdrawals	811	203	1 014

Table 3 - Number of observations per payment segment

Lastly, after computing the regressions prediction, a calibration study was developed. Conducting a calibration study involves comparing the predicted values generated by the regression model with the actual observed values of the target variable - in this case the average value of fraud per PSP and the share of fraud value/volume. Moreover, this method allows to determine whether the predictions exhibit biases or relevant discrepancies compared to actual

data. The calibration was held as a process used to assess the accuracy and reliability of predictions made by the models. However, in a second-layer reflection it could be also used as a method to better adjusts or improve the regressions and to enhance their predictive validity.

4.5. Evaluation metrics

Regression models are frequently evaluated using several metrics to estimate their accuracy and predictive power. In this investigation, the metrics took in consideration to evaluate the linear regression models were the following:

- **R-squared:** This coefficient of determination is a statistical measure valuable for assessing the goodness of fit of regression models and that quantifies the percentage of a regression model's independent variable that is explained by the predictors. This indicator typically ranges from 0 to 1, with higher values indicating a better fit on the data. Additionally, intermediate values indicate the percentage of the total variance that the model explains.

  R-squared might not give an overall picture of the performance of the model, particularly for regressions with several predictors. Furthermore, the metric does not indicate whether the model is overfitting the data or whether the predictions are made objectively. Therefore, must be interpreted combined with additional metrics.

  The formula to calculate this metric is the following:

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{y}_i)^2}{\sum_i (y_i - \bar{y})^2} \quad \text{or} \quad R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$

- **Adjusted R-squared:** This metric is an adjusted version of the coefficient of determination that adjusts for the number of predictors in a regression model. The adjusted measure is especially helpful for evaluating the general goodness of fit of multiple predictor models or comparing regression models with varying numbers of predictors.

$$R^2_{\text{adj}} = 1 - \frac{(1-R^2)\times(n-1)}{n-k-1}$$

- **Root Mean Squared Error (RMSE):** RMSE quantifies the average difference between the predicted and observed values, the lower results reflect a minor error magnitude between both values. This variable is scale-dependent, meaning that its value is influenced by the scale of the target variable and is expressed in the same units as the target variable, all the numeric variables were normalized in order to uniformize the

scales. Furthermore, is important to notice that this metric is sensitive to outliers and, if not properly handled, these unexpected values can potentially skew the assessment of the model's performance.

$$RMSE = \sqrt{\sum_{i=1}^{n} \frac{(\hat{y}_i - y_i)^2}{n}}$$

- **Mean Absolute Error (MAE):** MAE measures the average absolute difference between the predicted and observed values, meaning that, a lower error value indicates better performance. Regarding the units, similar to the RMSE they are expressed in the same units as the outcome variable.

  MAE is less sensitive to outliers due to the absolute computation in the formula. Therefore, this metrics complement a big missing gap in the evaluation metrics until now, the scenario where outliers could impact the model.

$$MAE = \frac{1}{n}\sum_{i=1}^{n} |y_i - \widehat{y_i}|$$

- **F-statistic:** Unlike the previous metrics that are part of the calibration accuracy test, this indicator measures the overall significance of the regression model by testing if the regression as a whole is statistically significant in explaining the variability in the target variable, essentially testing whether the addition of predictors significantly improves the model's fit to the data. Moreover, the F-test assumes a linear relationship between predictors and the dependent variable and follows an F-distribution, which is a right-skewed distribution with a range of values greater than or equal to zero. Furthermore, as the model's explanatory power increases, the F-statistic increases accordingly.

$$F = \frac{\frac{\text{SSR}}{k}}{\frac{\text{SSres}}{n-k-1}}$$

Notation	Description
y_i	Observed value for the *i*-th observation.
$\hat{y}_i$	Predicted value for the *i*-th observation.
SSRres	Residual Sum of Squares.
SSRtot	Total Sum of Squares.
SSR	Regression Sum of Squares

n	Number of observations.
k	Number of independent variables.

Table 4 – Formula's legend

5. Discussion and Results

In this section, the results of the linear regression models are presented as well as the calibration results.

The variables of interest, defined according to the analysis carried out, have been included in the linear models. However, the accuracy changes according to their relationship with *y*. The same model was predicted for different *y* metrics used to calculate fraud. lm1 corresponds to the model that predicts lm1 (i) the average value of fraud, lm2 (ii) the percentage of fraud in volume and lm3 (iii) the percentage of fraud in value. The primary aim is to address the research inquiries effectively.

As previously mentioned, the regressions were computed by two perspectives points, namely the issuer and the acquirer categorized into remote/non-remote transactions, additionally, the cash withdrawals were also included. Therefore, there are five optimal models, one per segment. Furthermore, for each model, graphical representations were generated to facilitate visualization of the calibration test. In these graphs, only the model with the best accuracy metrics for a given dependent variable is shown, along with its standard deviation.

5.1. Remote cards (acquirer)

Figure 23 present lm3 coefficients for the regression models assuming remote card fraud from the acquirer's perspective. The variables that had the greatest impact on the share of fraud value, according to the *p-value*, were the socio-economic indicators that were added to the regression, namely covid, education, inflation, and GDP.

By the analysis of the Figure 23, during the covid period, fraud is expected to decrease as well as with a 1 unit increase in GDP variation. This decline can be attributed to various factors, including economic uncertainty, shifts in consumer behaviour towards essential purchases, reinforced security concerns or, for instance, disruptions in supply chains. On the other hand, with a 1 unit increase in tertiary education variation or change in inflation variation, fraud is expected to increase. Tertiary education often correlates with higher incomes and a greater propensity for online shopping among educated individuals. On the other hand, inflationary pressures may drive consumers to seek alternative payment options, such as remote card

payments, to mitigate the impact of rising prices. Both indicators lead to an increase of transactions and consequently fraud.

	estimate	std.error	statistic	p.value
(Intercept)	-2.451553e-05	3.565481e-06	-6.87580055	2.248320e-11
Type_fraud_407	-2.298021e-07	4.520966e-07	-0.50830313	6.115083e-01
SCA	1.825874e-07	1.954345e-07	0.93426404	3.507058e-01
Motive_N_SCA_6	2.328014e-09	2.319050e-07	0.01003865	9.919952e-01
Motive_N_SCA_9	-2.772381e-08	2.172771e-07	-0.12759655	8.985295e-01
Motive_N_SCA_10	6.103398e-08	1.784654e-07	0.34199328	7.325274e-01
Motive_N_SCA_11	5.721540e-08	2.629790e-07	0.21756644	8.278728e-01
covid	-1.564353e-06	2.209365e-07	-7.08055511	6.095476e-12
ukr_war	-3.665687e-07	2.712915e-07	-1.35119900	1.773608e-01
edu_pt	1.065085e-04	1.381218e-05	7.71120246	9.163504e-14
inf_pt	2.893966e-01	3.892445e-02	7.43482873	5.956678e-13
gdp_pt	-2.149422e-01	3.308556e-02	-6.49655655	2.334168e-10
nov_dec:dec_jan	1.421582e-07	1.882132e-07	0.75530406	4.504909e-01

Figure 20 - Remote Cards (Acquirer) best model Coefficients

Table 5 presents relative accuracy metrics from the same segment. Here we can here remark that the model with the best results is the lm3 model where the dependent variable is the share of fraud value and the variables considered are the ones presented in Figure 23. The lower values of RMSE (26%) and MAE (36%) indicate better model performance compared to the other models, as they reflect smaller deviations between the predicted and observed values. The R-squared and Adjusted R-squared value of approximately 0.5 units suggests that 50% of the variance in the dependent variable is explained by the independent variables.

Relative Accuracy Metrics

Model	MAE	RMSE	R_squared	Adjusted_Rsquared	F_statistic
lm1	0.814	1.264	0.041	0.014	1.510
lm2	0.333	0.499	0.166	0.142	6.992
lm3	0.265	0.355	0.515	0.501	37.106

Table 5 – Remote Cards (Acquirer) evaluation metrics

Figure 24 illustrates the actual values compared to those predicted by the lm3 model. Both original and lm3 observations have distinct patterns, in fact, the original observations are characterized by peaks and high variance. While this pattern sets challenges for computation, it is noteworthy that the regression standard deviation aligns closely with the majority of observations, despite accurately predicting only a subset of them.

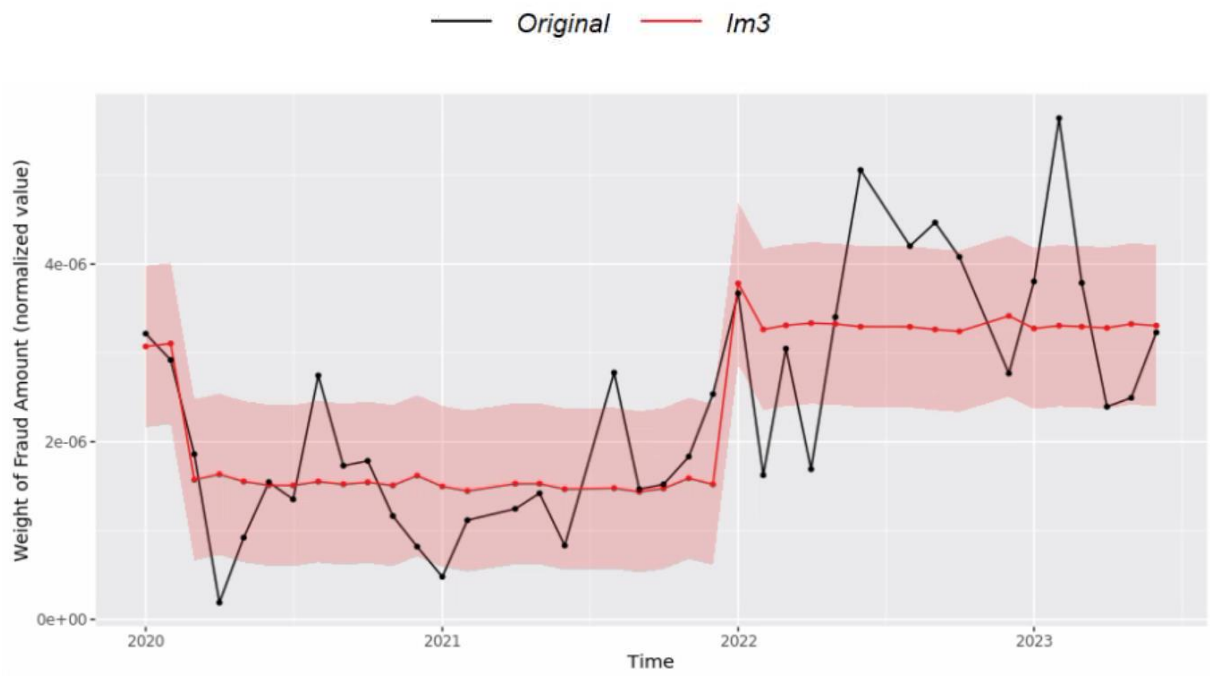

Figure 21 - Remote Cards (Acquirer) Calibration between the predicted and original observations

5.2. Non-remote cards (acquirer)

In the non-remote segment, the cards (acquirer) exhibit externally additional variables to the model as significant. Figure 25 denotes lm3 coefficients for the regression models. Similar to before, the covid-19 pandemic has a negative impact, while all education, inflation, and GDP variation show positive impacts. It is noteworthy that although not having a *p-value* lower than 0.05, the variable representing the impact of the Ukraine war has a relatively low *p-value*. During periods of geopolitical uncertainty or conflict, consumers may prefer card payments over cash transactions due to their perceived safety and convenience, particularly for larger purchases or transactions. This preference for card payments could result in an increase in the volume and value of non-remote card transactions.

The variable Type_fraud_407 also has a significant impact on fraud, a 1 unit increase in the modification of a payment order by the fraudster result in a decrease on average in the share of fraud value. In these regressions, the SCA component is not included, as expected in non-remote fraud, hence the variable Motive_N_SCA_10 being highly significant, as it relates to the fact that these operations are not subject to SCA.

	estimate	std.error	statistic	p.value
(Intercept)	-2.261222e-05	3.316099e-06	-6.81892241	1.660495e-11
Type_fraud_402	-2.850418e-07	1.015687e-07	-2.80639525	5.115818e-03
Type_fraud_407	-1.012843e-08	1.505917e-07	-0.06725753	9.463913e-01
Motive_N_SCA_7	7.495526e-08	1.851714e-07	0.40478860	6.857271e-01
Motive_N_SCA_8	-5.800671e-08	1.788086e-07	-0.32440671	7.457040e-01
Motive_N_SCA_9	-1.603735e-08	1.623355e-07	-0.09879136	9.213255e-01
Motive_N_SCA_10	-1.107033e-06	3.272516e-07	-3.38282081	7.477521e-04
Motive_N_SCA_255	8.289857e-08	1.624215e-07	0.51039150	6.098998e-01
Channel_2	-2.068051e-06	6.850757e-07	-3.01871846	2.608551e-03
Channel_4	-1.950573e-06	5.400932e-07	-3.61154845	3.208540e-04
Channel_5	-1.927357e-06	5.588492e-07	-3.44879715	5.886701e-04
covid	-1.457166e-06	1.845682e-07	-7.89499803	8.246571e-15
ukr_war	4.555445e-07	2.541449e-07	-1.79246015	7.338811e-02
edu_pt	1.069275e-04	1.274496e-05	8.38978743	1.821979e-16
inf_pt	2.729411e-01	3.534677e-02	7.72181141	2.990693e-14
gdp_pt	2.054259e-01	2.973475e-02	-6.90861422	9.135525e-12
nov_dec:dec_jan	-1.083812e-07	1.202271e-07	-0.90147049	3.675744e-01

Figure 22 - Non-remote Cards (Acquirer) best model Coefficients

Table 6 compares the same model to different dependent variables. Concerning non-remote card fraud from the acquirer's perspective, the lm3 model stands out. Similar to the remote scenario, the accuracy metrics collecting the most attention are those from the model with share of fraud value as the target variable. The errors, as measured by the RMSE and MAE metrics, are relatively low 36% and 25%, respectively, indicating that, on average, the model predictions exhibit minimal deviation from the actual values.

Relative Accuracy Metrics

Model	MAE	RMSE	R_squared	Adjusted_Rsquared	F_statistic
lm1	1.159	1.960	0.104	0.089	6.725
lm2	0.364	0.514	0.208	0.194	15.121
lm3	0.254	0.357	0.427	0.417	42.872

Table 6 - Non-remote Cards (Acquirer) evaluation metrics

Figure 26 that compare the observed values to the predictions generated by the lm3 model, reflects a model that does not objectively predict the observations, despite the previous positive coefficient results and metrics. Overall, the model has some pattern similarities with lm3 in Figure 24.

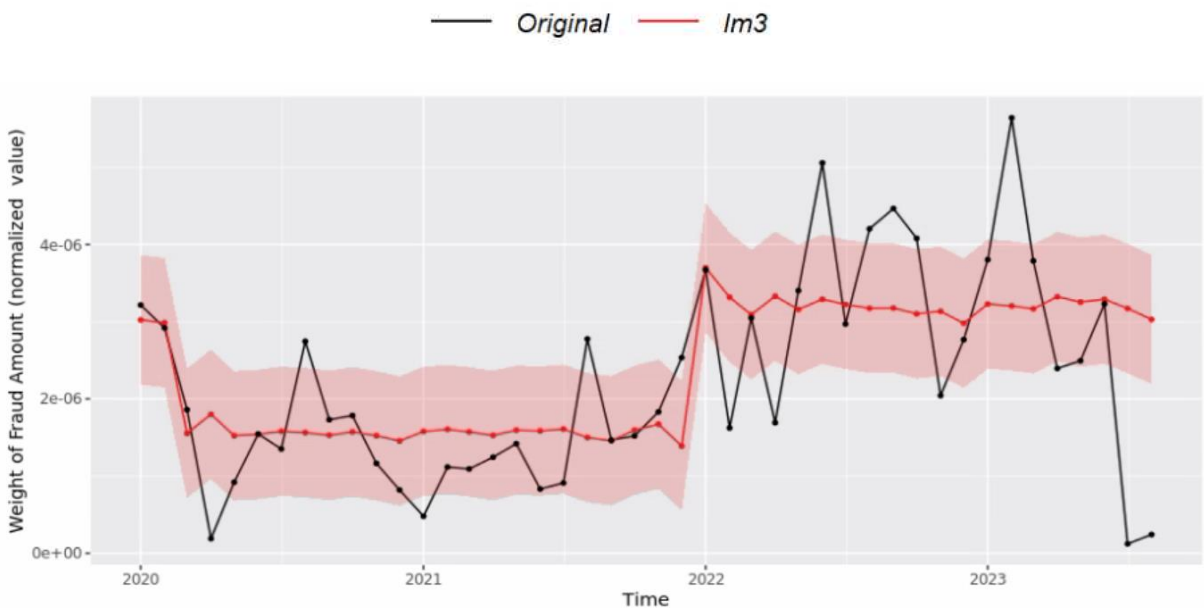

Figure 23 - Non-remote Cards (Acquirer) Calibration between the predicted and original observations

5.3. Remote cards (issuer)

Figure 25 exhibits lm3 coefficients for the regression models of the remote card fraud in the issuer perspective. The seasonal component between November and January, included in the interaction terms nov_dec:dec_jan, aims to understand the impact of events such as Christmas, New Year's, and sales on fraud. The coefficient suggests a significant negative relationship between these time periods and fraud occurrence, indicating a reduction in the share of fraud value in remote cards when these events coincide. Additionally, significant variables, such as inflation, tertiary education level, and unemployment, suggest potential influences on fraud patterns. High unemployment periods may lead individuals to engage in illicit activities or accept risky financial offers. On the other hand, as the model suggests, since unemployment has a negative coefficient, during such periods, individuals may reduce discretionary spending, including online purchases, decreasing the share of fraud value.

	estimate	std.error	statistic	p.value
(Intercept)	-1.213758e-05	4.104271e-06	-2.95730449	3.108228e-03
Type_fraud_402	1.118310e-06	6.064976e-07	1.84388197	6.521996e-02
Type_fraud_407	2.244310e-07	1.156622e-06	0.19404011	8.461471e-01
SCA	1.390921e-06	6.301353e-07	2.20733624	2.730559e-02
Motive_N_SCA_3	5.822286e-07	4.057636e-06	0.14348960	8.859054e-01
Motive_N_SCA_4	1.597514e-07	8.129310e-07	0.19651284	8.442114e-01
Motive_N_SCA_6	6.700288e-08	6.791591e-07	0.09865565	9.214130e-01
Motive_N_SCA_9	1.888645e-06	5.857671e-07	3.22422508	1.265830e-03
Motive_N_SCA_10	1.644841e-07	5.852021e-07	0.28107223	7.786589e-01
Motive_N_SCA_11	9.178416e-07	2.256169e-06	0.40681415	6.841503e-01
Motive_N_SCA_12	-1.974757e-06	2.654775e-06	-0.74385094	4.569783e-01
Channel_3	-1.548293e-06	7.618779e-07	-2.03220585	4.215032e-02
Channel_5	-9.141736e-07	2.607678e-07	-3.50569956	4.567322e-04
covid	-2.069249e-05	4.739479e-07	-43.65984342	0.000000e+00
ukr_war	3.494100e-06	6.654710e-07	5.25056677	1.536913e-07
unemp_pt	-2.000847e+00	1.598478e-01	-12.51719749	9.054481e-36
edu_pt	1.954421e-04	1.575320e-05	12.40650480	3.577717e-35
inf_pt	1.906065e+00	9.066548e-02	21.02305354	9.823696e-97
nov_dec:dec_jan	-2.899755e-06	4.156418e-07	-6.97657253	3.150651e-12

Figure 24 - Remote Cards (Issuer) best model Coefficients

In Table 7 a comparison between models with different *y* metrics is made. Lm3 that compute the share of fraud value is, arguably, the optimal model in the remote issuer's perspective. Despite the lm2 model has a lower MAE error value, the model with greater statistical significance and better explanatory power for the variation of the dependent variable is lm3.

Relative Accuracy Metrics

Model	MAE	RMSE	R_squared	Adjusted_Rsquared	F_statistic
lm1	1.066	2.691	0.025	0.024	22.116
lm2	0.320	0.622	0.113	0.112	107.012
lm3	0.343	0.628	0.143	0.142	140.875

Table 7 – Remote Cards (Issuer) evaluation metrics

The graphical representation of lm3 in Figure 28 does not capture the peaks observed in the first semester of 2021 and 2022. During this period, the model predicts less accurately, as the observations exhibit greater variability.

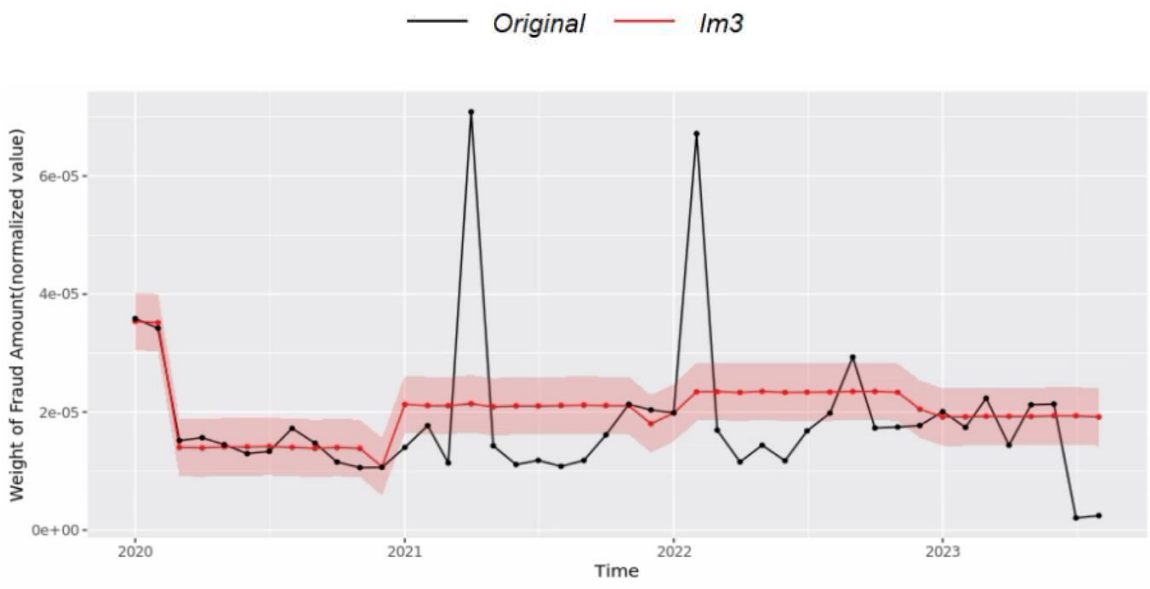

Figure 25 - Remote cards (issuer) Calibration between the predicted and original observations

5.4. Non-remote cards (issuer)

Figure 29 presents the lm2 coefficients for the regression model of the non-remote card issuer segment. The type of fraud "Modification of a payment order by the fraudster" and "Manipulation of the payer by the fraudster to issue a payment order", have a significant impact, respectively negative and positive on the share of fraud in volume. It is also important to note the inverse relationship between fraud and the level of unemployment variation and with the types of channels (debit, credit, and delayed debit cards).

	estimate	std.error	statistic	p.value
(Intercept)	-9.705257e-05	2.163432e-05	-4.4860478	7.330701e-06
Type_fraud_402	-5.749951e-06	1.364749e-06	-4.2131940	2.538328e-05
Type_fraud_407	2.073750e-05	5.918681e-06	3.5037362	4.606393e-04
Motive_N_SCA_3	-3.828024e-06	2.381407e-05	-0.1607463	8.722962e-01
Motive_N_SCA_4	8.464940e-06	1.784101e-05	0.4744654	6.351777e-01
Motive_N_SCA_7	1.141597e-05	3.322760e-06	3.4356893	5.932753e-04
Motive_N_SCA_8	7.228568e-08	4.830316e-06	0.0149650	9.880604e-01
Motive_N_SCA_9	1.299592e-05	2.825380e-06	4.5997071	4.279039e-06
Motive_N_SCA_10	9.064308e-06	2.957547e-06	3.0648063	2.183490e-03
Motive_N_SCA_12	-1.456185e-05	5.296546e-05	-0.2749311	7.833744e-01
Motive_N_SCA_255	3.021973e-05	2.986151e-06	10.1199586	5.764419e-24
Channel_2	-3.092137e-05	4.344161e-06	-7.1179158	1.165371e-12
Channel_4	-1.009276e-05	2.633282e-06	-3.8327679	1.274291e-04
Channel_5	-4.929036e-06	3.859256e-06	-1.2771986	2.015596e-01
covid	-3.147236e-06	2.430662e-06	-1.2948064	1.954147e-01
ukr_war	4.276257e-05	3.634571e-06	11.7655088	9.187053e-32
unemp_pt	-1.760425e+01	8.479712e-01	-20.7604382	6.670389e-94
edu_pt	7.585415e-04	8.095142e-05	9.3703305	8.671807e-21
inf_pt	1.262553e+01	4.477819e-01	28.1957094	7.957132e-169
nov_dec:dec_jan	-2.165838e-05	2.077564e-06	-10.4248932	2.517748e-25

Figure 26 - Non-remote Cards (Issuer) best model Coefficients

From Table 8 we can state that for non-remote cards from the issuer's perspective, the model with the highest accuracy results is lm2. This model measures the share of fraud volume. The RMSE error values may be considered high, unlike the R squared and adjusted R squared which indicate that a significant portion of the variability in the dependent variable is explained by the independent variables.

Relative Accuracy Metrics

Model	MAE	RMSE	R_squared	Adjusted_Rsquared	F_statistic
lm1	1.073	1.173	0.034	0.032	20.341
lm2	0.377	0.638	0.142	0.141	94.605
lm3	0.390	0.645	0.164	0.158	29.287

Table 8 - Non-remote Cards (Issuer) evaluation metrics

Figure 30 shows that the lm2 model accurately predicted the actual results in the first year of analysis and in 2023, however it did not follow the evolution between 2021 and 2022 mostly due to an increase in the variance of the observations.

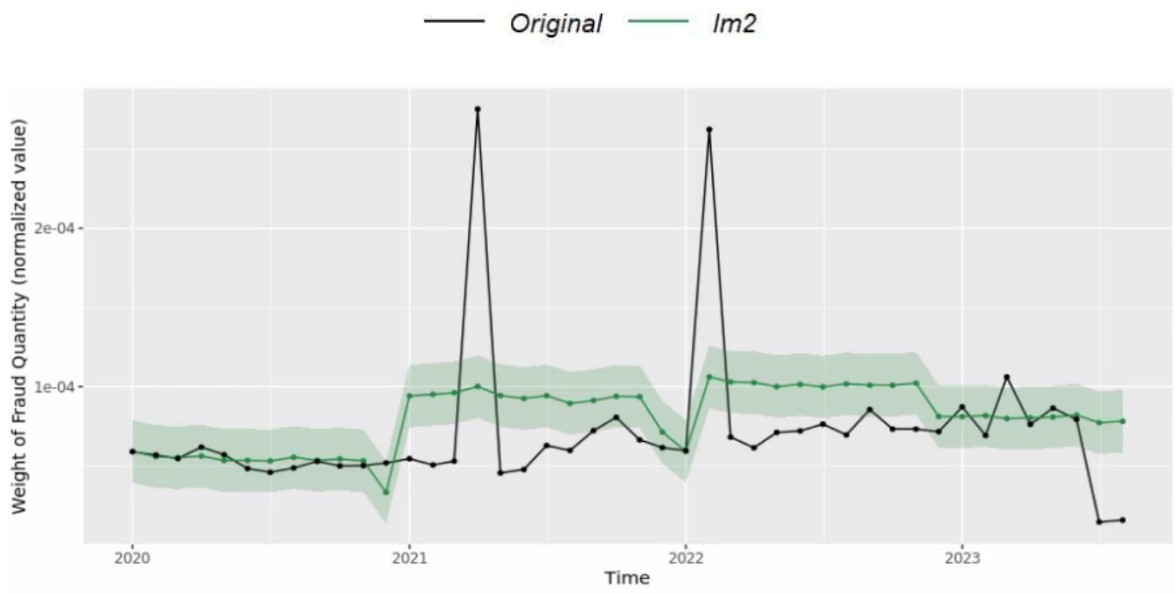

Figure 27 - Non-remote Cards (Issuer) Calibration between the predicted and original observations

5.5. Cash Withdrawals

Figure 31 shows the lm3 coefficients of the cash withdrawals segment. In this model a new variable internet_pt was introduced. In fact, an increase of 1 unit of the private households with a computer internet access, and broadband internet connection decrease in the share of fraud value in cash withdrawals. In addition, e-commerce, and payment online increase due to easier access to the internet. Covid plays a crucial role in the decrease of fraud via withdrawals since the payments become more electronic.

	estimate	std.error	statistic	p.value
(Intercept)	8.648307e-06	1.926094e-06	4.4900753	1.223278e-05
Type_fraud_402	-5.566108e-08	1.272348e-07	-0.4374672	6.622617e-01
Type_fraud_407	9.817545e-08	8.184812e-08	1.1994833	2.318105e-01
Funtion_Card_2	-3.742744e-08	3.577547e-08	-1.0461758	2.967883e-01
covid	-2.928569e-06	7.581352e-08	-38.6285932	9.331037e-93
ukr_war	1.955858e-07	1.299220e-07	1.5054098	1.338536e-01
unemp_pt	-4.417533e-02	3.324896e-02	-1.3286229	1.855412e-01
internet_pt	-8.147692e-02	3.123790e-02	-2.6082715	9.812208e-03
inf_pt	4.016710e-02	1.746035e-02	2.3004743	2.249034e-02
nov_dec:dec_jan	-6.787169e-08	7.076770e-08	-0.9590771	3.387198e-01

Figure 28 - Cash withdrawals best model Coefficients

Within the cash withdrawals payment instrument the optimal model is lm3 as demonstrated in Table 9. In both RMSE and MAE metrics the error is relatively low, 25% and 17%, respectively.

Moreover, the R squared and Adjusted R squared are close to one indicating that around 91% of dependent variable is explained by the covariates.

Relative Accuracy Metrics

Model	MAE	RMSE	R_squared	Adjusted_Rsquared	F_statistic
lm1	0.4663574	0.8199580	0.0449337	0.0342026	4.187248
lm2	0.2239177	0.3442821	0.7286653	0.7256166	239.008186
lm3	0.1659638	0.2541756	0.9120715	0.9079713	222.440695

Table 9 - Cash Withdrawals evaluation metrics

The graphical representation in Figure 32, shows that lm3 accurately predicted the observation through the years of the analysis, since the actual observations are included in the variance margin of the model.

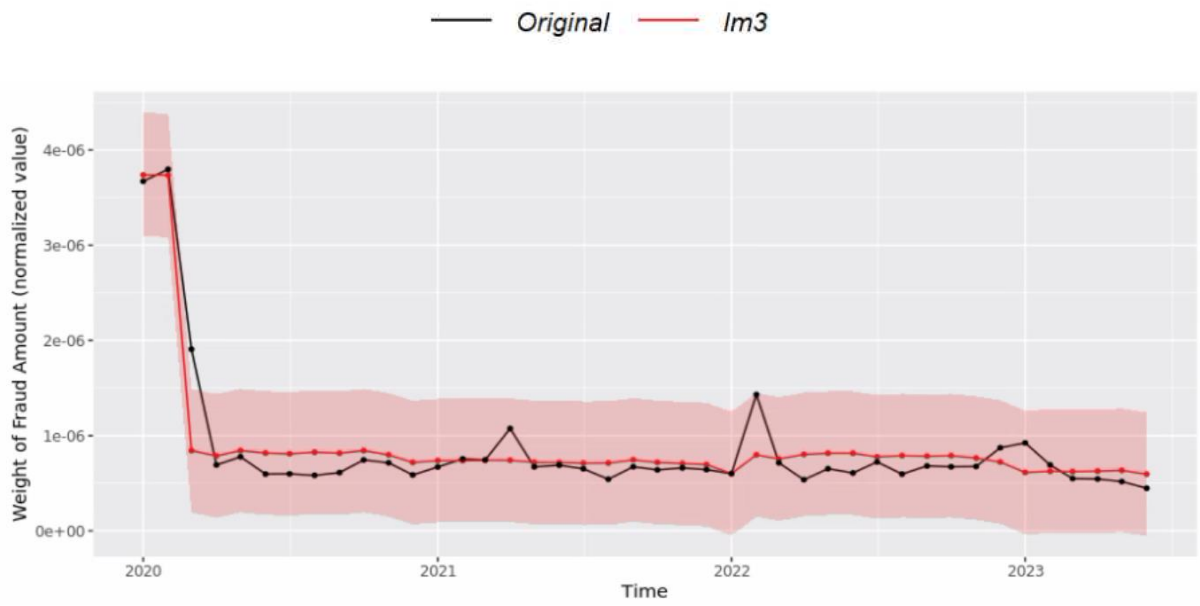

Figure 29 - Cash Withdrawals Calibration between the predicted and original observations

Based on the analysis conducted, there is compelling evidence to reject both H0-1 and H0-2 hypotheses proposed. Despite the results not being optimal, the variables that most impacted fraud were identified. The regression analyses conducted revealed that payment transactions characteristics and economic conditions display significant associations with card fraud. Moreover, the obtained *p-values* for each variable were statistically significant.

In general, the results of the linear models did not exactly predict the results of the *y* variable mainly due to the variance and unpredictability of fraud. However, in most cases, the actual observations are within the range of the standard deviation.

6. Limitations

After analysing the results, it is essential to acknowledge the limitations of this research. The findings of this study are focused on fraud within card payments and cash withdrawals. As expected, each subgroup must be analysed individually. The segmentation of the study into remote and non-remote transactions intend to mitigate potential errors resulting from the consideration of distinct characteristics within a single regression analysis. Therefore, the data should ideally be evenly distributed across the five regressions, which is not the case. The number of observations in some segments might be insufficient, namely Cards (acquirer) and Cash Withdrawals.

Secondly, the analysis was conducted using data from 2020 onwards, comprising a sample of three years. This temporal scope may not be extensive enough to draw significant conclusions regarding temporal patterns. Additionally, economic variables typically exhibit gradual variations over time. A longer time span than three years is preferable when capturing complete economic cycles.

Some values were challenging to filter as either accurate or misaligned with reality. In fact, linear regression is susceptible to outliers and influential points, which can impact regression coefficients and predictions.

In this analysis, ensemble learning methods such as Gradient Boosting Machines and Random Forests could have been employed instead of the linear regression. Additionally, time series methods such as ARIMA could have been explored in the prediction section.

Lastly, multicollinearity can also be a limitation in this investigation. By having two or more variables highly correlated it can lead to unstable coefficient estimates, making it difficult to determine the true effect of each independent variable on y. Despite trying to remove from the models the highly correlated relationships, there might still be some that could lead the model to overfit.

7. Conclusion

This study aims to contribute to the identification of broad factors that may impact card fraud. The main goal is to explore the significance of specific attributes in card fraud and gather insights on the subject.

In remote cards, both acquirer and issuer perspectives have covariates that significantly influence fraud. Firstly, is important to refer that covid-19 impacted negatively remote card fraud. This might possibly be influenced by increased awareness and vigilance among consumers regarding online transactions, as people focus their transaction on specific sectors and became more cautious about potential fraud risks associated with remote payments. Moreover, during the covid-19 pandemic, cash withdrawals decreased, as the population was in lockdown and not directly consuming at physical stores.

The second major variable affecting negatively remote card fraud is GDP variation. Reduced consumer willingness to pay during economic downturns may also result in decreased online shopping and less willingness to engage in high-value transactions. Furthermore, in remote Cards (issuer perspective specifically), there is a negative impact on fraud when considering the Mobile payment solution channel, possibly due to the preference for other initiation channels.

On the other hand, inflation and education variation have a positive and significant impact on remote card fraud. The potential for inflation to diminish consumer purchasing power could drive individuals towards alternative payment methods, including online transactions facilitated by credit or debit cards. This shift towards increased reliance on card-based payments might create additional opportunities for exploitation by fraudsters.

For the non-remote segment, the variables that most impacted fraud, in this case negatively, are the type of fraud - “Modification of a payment order by the fraudster” as well as all three types of channels (ATM or other PSP terminal and Initiated at a physical EFTPOS). The behaviour of macro variables is similar to the remote segment, with the difference lying in the impact of GDP variation on fraud, which is positive in non-remote cards. This pattern may be attributed to a robust economy indicating higher consumer spending and confidence that could incentivize fraudsters.

Regarding the Ukraine military conflict, it is anticipated that its repercussions will become increasingly pronounced from 2023 onwards. Nevertheless, its effects have already manifested as an influence factor on both remote and non-remote card fraud.

In addition, the variable incorporating the digitization component in the models, impacted fraud in cash withdrawals. With increased digitization and, in this case, internet access, there is greater ease and exposure to online payments, which in a short-term can be quickly replace withdrawals.

The impact of SCA on fraud in remote cards was not significantly substantial in this analysis, as expected, due to its recent introduction. Therefore, in subsequent investigations, it would be pertinent to analyse the impact of SCA on both electronic and non-electronic card environments.

Moreover, the emphasis in a forthcoming analysis should not completely focus on the incorporation of different variables, but rather understand and justify in more detail the reason behind the impact of this macroeconomic variables on cards.

8. References

- Abrazhevich, Dennis. 2001. "Classification and Characteristics of Electronic Payment Systems". *IPO, Center for User-System Interaction Technical University of Eindhoven (TUE)*. Pages 81- 90.
- Ahmad, Bashir, Maria Ciupac-Ulici, and Daniela-Georgeta Beju. 2021. "Economic and Non-Economic Variables Affecting Fraud in European Countries". Risks 9: 119. https://doi.org/10.3390/risks9060119.
- Dal Pozzolo, Andrea, Olivier Caelen, Yann-Aël Le Borgne, Serge Waterschoot, Gianluca Bontempi. 2014. "Learned lessons in credit card fraud detection from a practitioner perspective". Université Libre de Bruxelles, Brussels, Belgium and Worldline, Brussels, Belgium.
- Delamaire, Linda, Hussein Abdou, John Pointon. 2009. "Credit card fraud and detection techniques: a review". Banks and Bank Systems, Volume 4, Issue 2.
- Dimitras, Augustinos I., Maria I. Kyriakou and George Iatridis. 2015. "Financial crisis, GDP variation and earnings management in Europe". Research in International Business and Finance. Vol. 34. pp 338-354.
- Edge, Michael Edward and Pedro R. Falcone Sampaio. 2009. "A survey of signature based methods for financial fraud detection". Computers & security 28. pp 381–394.
- Fernandes, Lina. 2013. "Fraud In Electronic Payment Transactions: Threats And Countermeasures". Asia Pacific Journal of Marketing & Management Review. Vol. 2. Pp 23-32.
- Final Report of the European Banking Authority (EBA/GL/2018/05). Guidelines on fraud reporting under the Payment Services Directive 2 (PSD2). Article 96(6). 18 July 2018. 141 pages·https://www.eba.europa.eu/sites/default/files/document_library/Guidelines%20amending%20EBA%20GL%20on%20Fraud%20reporting%20under%20PSD2.pdf.
- Hayashi, Fumiko. 2020. "Payment Card Fraud Rates in the United States Relative to Other Countries after Migrating to Chip Cards". Federal Reserve Bank of Kansas City. Pp 23-40.
- Hayes, Andrew F. and Li Cai. 2007. "Using heteroskedasticity-consistent standard error estimators in OLS regression: An introduction and software implementation". Behaviour Research Methods. Vol. 39 (4), pages 709-722.
- Hjalmarsson, Randi; Lance Lochner. 2012. "The Impact of Education on Crime: International Evidence", CESifo DICE Report, ISSN 1613-6373, ifo Institut - Leibniz-

Institut für Wirtschaftsforschung an der Universität München, München, Vol. 10, Iss. 2, pp. 49-55.

- Itoo, Fayaz, Meenakshi, Satwinder Singh. 2020. "Comparison and analysis of logistic regression, Naı¨ve Bayes and KNN machine learning algorithms for credit card fraud detection". https://doi.org/10.1007/s41870-020-00430-y.
- Li, Jundong, Kewei Cheng, Suhang Wang, Fred Morstatter, Robert P. Trevino, Jiliang Tang, and Huan Liu. 2017. "Feature Selection: A Data Perspective". ACM Comput. Surv. 50, 6, Article 94 (December 2017), 45 pages. https://doi.org/10.1145/3136625.
- Karpoff, Jonathan M. 2021. "The future of financial fraud". *Journal of Corporate Finance 66.*
- Nicolini, Gianni, Lucia Leonelli.2021." Financial Frauds on Payment Cards: The Role of Financial Literacy and Financial Education". *The International Review of Financial Consumers*, Volume.6 Issue.1. https://doi.org/10.36544/irfc.2021.6-1.1
- Paul, Pongku Kumar. 2020. "Strong Customer Authentication: Security Issues and Solution". University Of Turku Department of Future Technologies. pages 58.
- Peja, Dijora. 2023. "Predicting hourly demand for shared bicycles with weather data and machine learning models". *Dissertation submitted in Business Analytics at the Universidade Católica Portuguesa*.
- Raj, S. Benson Edwin, A. Annie Portia.2011. "Analysis on Credit Card Fraud Detection Methods". Department of CSE Karunya University, Coimbatore.
- Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation) [2016] OJ L 119/1.
- Regulation (EU) 1409/2013 on payments statistics (ECB/2013/43) (ECB/2020/59).
- Rodríguez, Pau, Miguel A. Bautista, Jordi Gonzàlez and Sergio Escalera. 2018. "Beyond one-hot encoding: Lower dimensional target embedding". Image and Vision Computing. Vol 75. Pp 21-31. https://doi.org/10.1016/j.imavis.2018.04.004.
- Sahin, Yusuf, Serol Bulkan, Ekrem Duman. 2013. "A cost-sensitive decision tree approach for fraud detection". Marmara University, Kadikoy, 34722 Istanbul, Turkey, and Ozyegin University, Cekmekoy, 34794 Istanbul, Turkey.
- Saranya, C. and G.Manikandan. 2013. "A Study on Normalization Techniques for Privacy Preserving Data Mining". *International Journal of Engineering and Technology (IJET). Vol 5 No 3. Pages 2701-2704.*

https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=35a87b51f7441a87adee91e12eb4d22cd2565556.

- Kondo, Satoshi, Daisuke Miyakawa, Kengo Shiraki, Miki Suga, And Teppei Usuki, 2019, "Using Machine Learning to Detect and Forecast Accounting Fraud". *The Research Institute of Economy, Trade and Industry*.
- Sakharova, Irina. 2012. "Payment Card Fraud: Challenges and Solutions". The University of Texas at Dallas, Richardson, Texas, USA.
- Sharma, Anuj and Prabin Kumar Panigrahi. 2012. "A Review of Financial Accounting Fraud Detection based on Data Mining Techniques". *International Journal of Computer Applications*. Vol. 39. No.1. pages 37 – 47.
- Sharma, Anupam. 2020. "Changing Consumer Behaviours Towards Online Shopping - An Impact Of Covid 19". School of Humanities and Social Sciences, Thapar Institute of Engineering and Technology, Punjab, India, Volume 24, Issue 3.
- Smiti, Abir.2020. "A critical overview of outlier detection methods". LARODEC, University of Tunis, Tunísia Institut Supérieur de Gestion de Tunis.
- Sulaiman, Rejwan Bin, Vitaly Schetinin, Paul Sant. 2022." Review of Machine Learning Approach on Credit Card Fraud Detection". https://doi.org/10.1007/s44230-022-00004-0.
- Titalessy, Pisi Bethania. 2020. "Cashless Payments and its Impact on Inflation". *Advances in Social Sciences Research Journal*, 7. pp 524-532.
- Varmedja, Dejan, Mirjana Karanovic, Srdjan Sladojevic, Marko Arsenovic, Andras Anderla. 2019. "Credit Card Fraud Detection - Machine Learning methods". University of Novi Sad Novi Sad, Serbia.
- Wooldridge, Jeffrey M. 2012. "Introductory Econometrics: A Modern Approach", Fifth Edition. *Michigan State University. South-Western, Cengage Learning.*
- Zandi, Mark, Virendra Singh, Justin Irving.2013. "The Impact of Electronic Payments on Economic Growth". Moody's Analytics.

A. Appendix

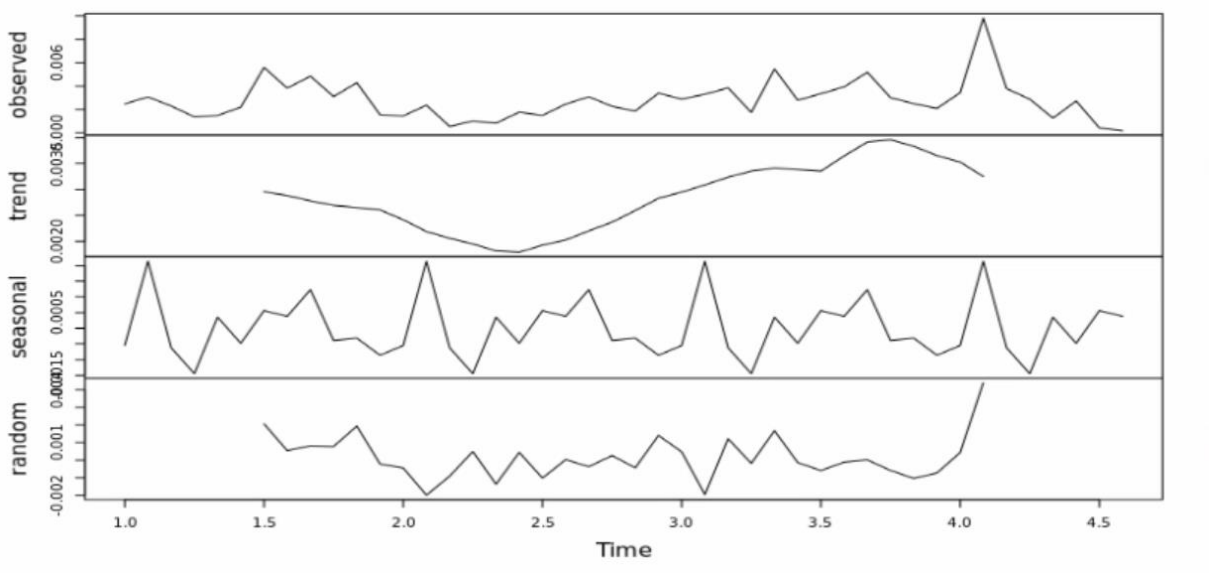

Figure 30 - Remote Cards Acquirer Time Decomposition

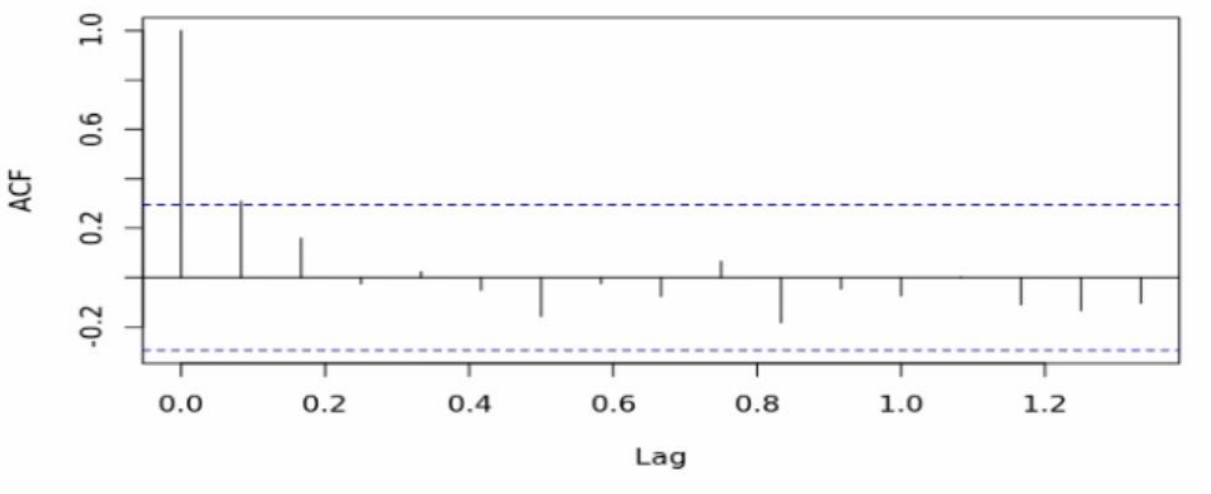

Figure 31 - Remote Cards Acquirer ACF

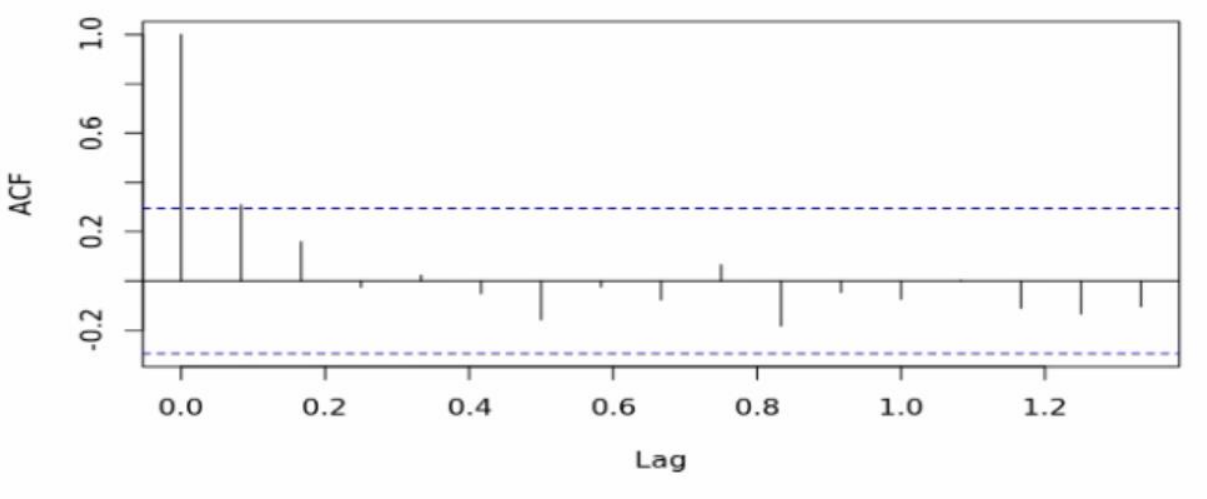

Figure 32 - Remote Cards Acquirer PACF

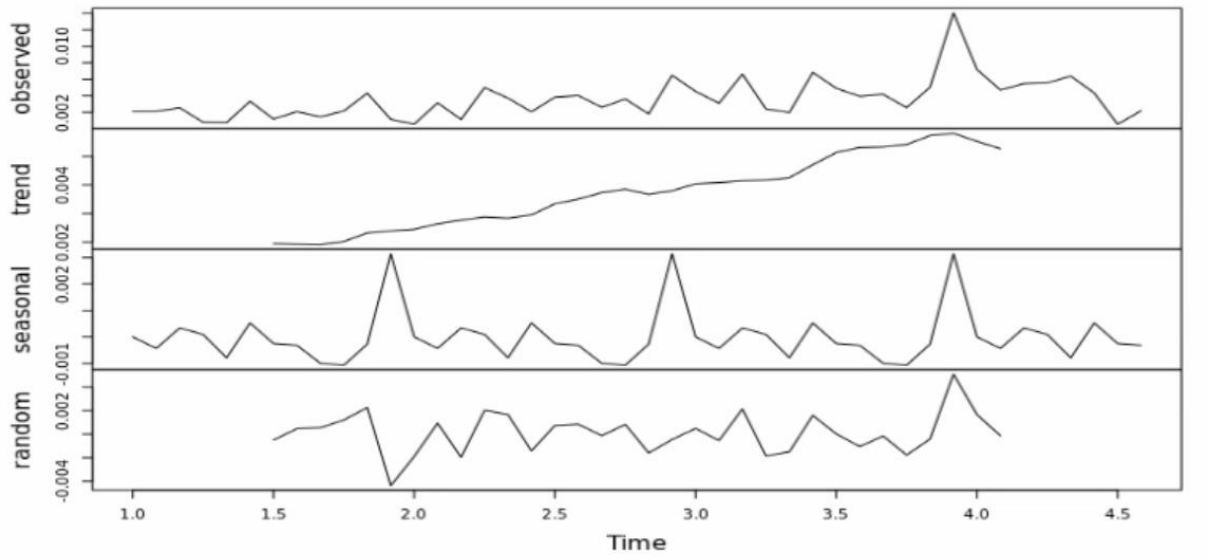

Figure 33 - Non-Remote Cards Acquirer Time Decomposition

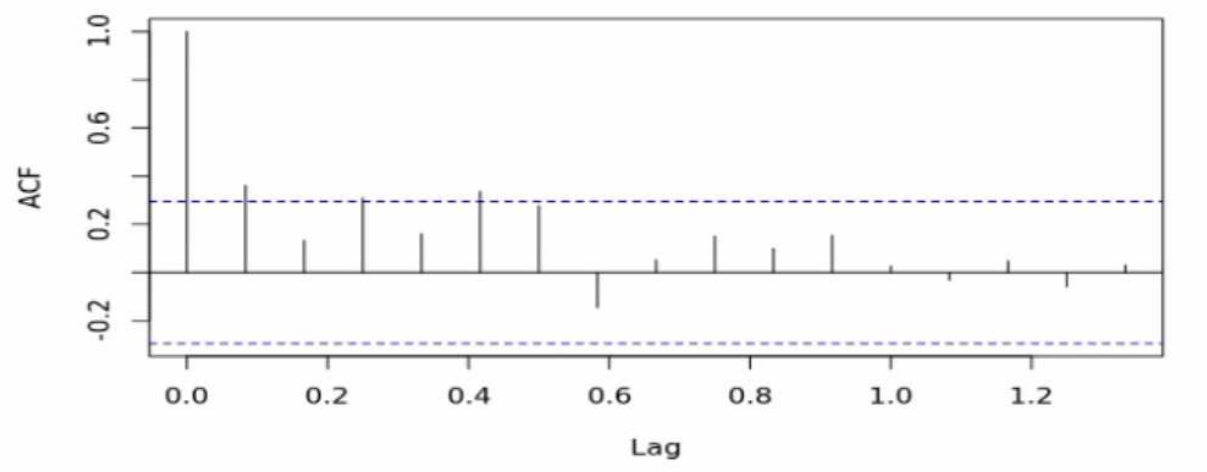

Figure 34 - Non-Remote Cards Acquirer ACF

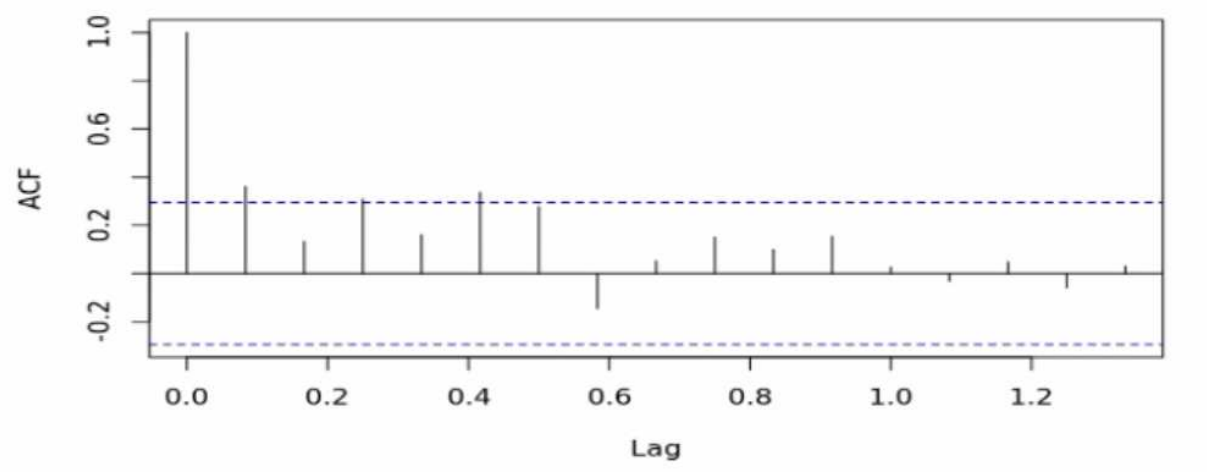

Figure 35 - Non-Remote Cards Acquirer PACF

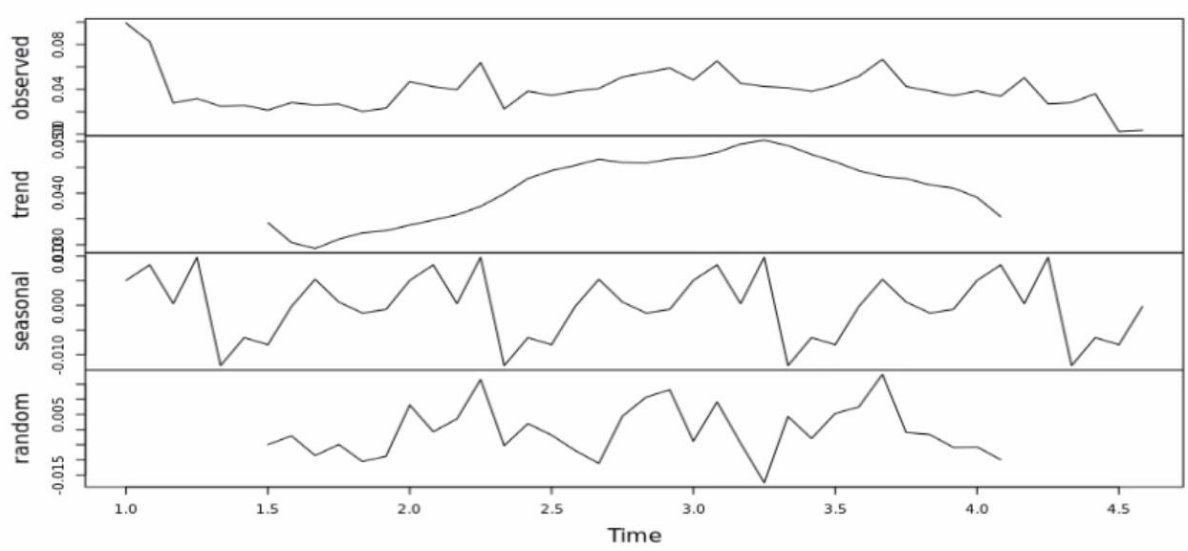

Figure 36 - Remote Cards Issuer Time Decomposition

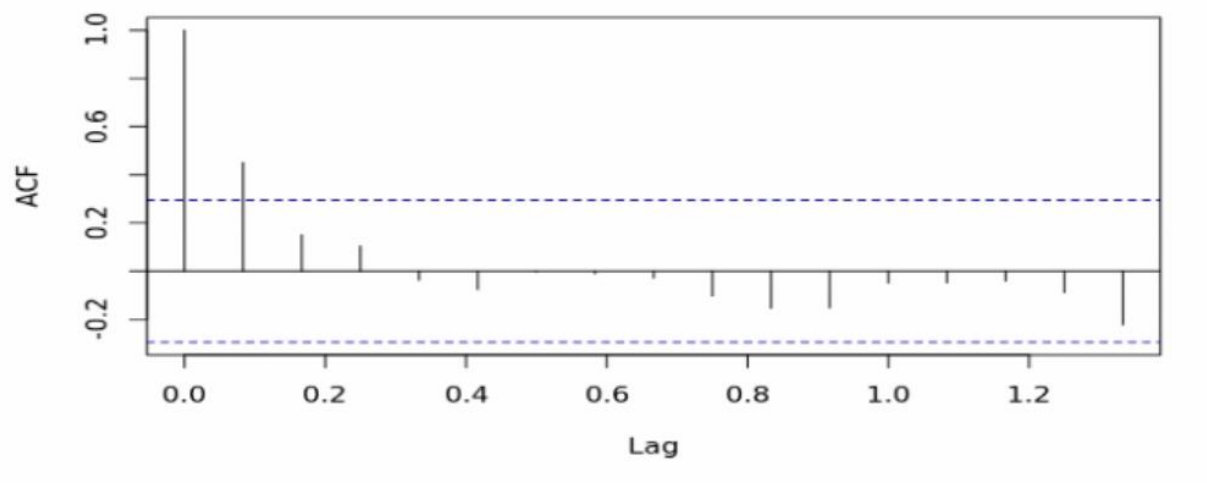

Figure 37 - Remote Cards Issuer ACF

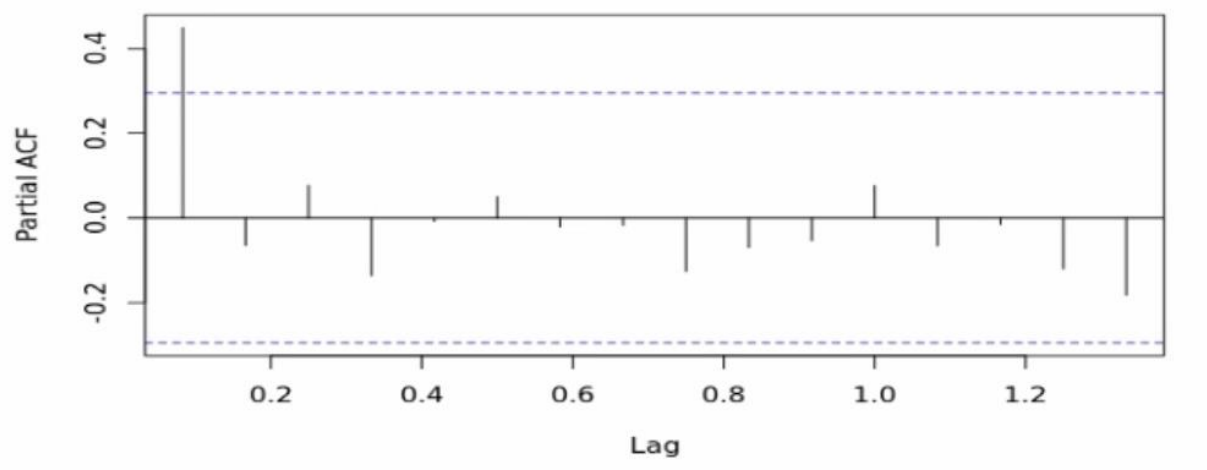

Figure 38 - Remote Cards Issuer PACF

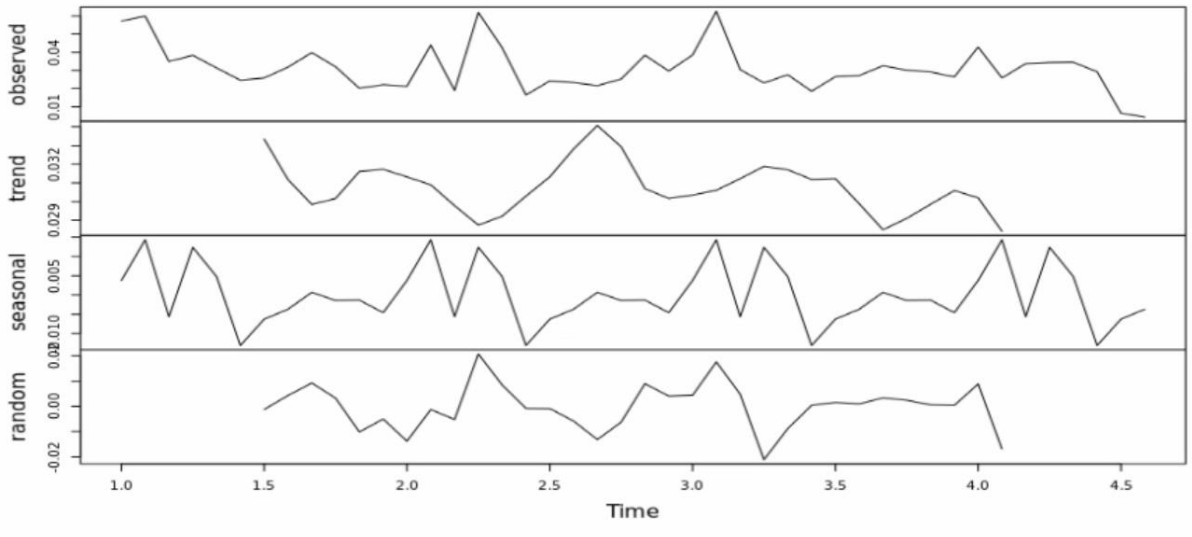

Figure 39 - Non-Remote Cards Issuer Time Decomposition

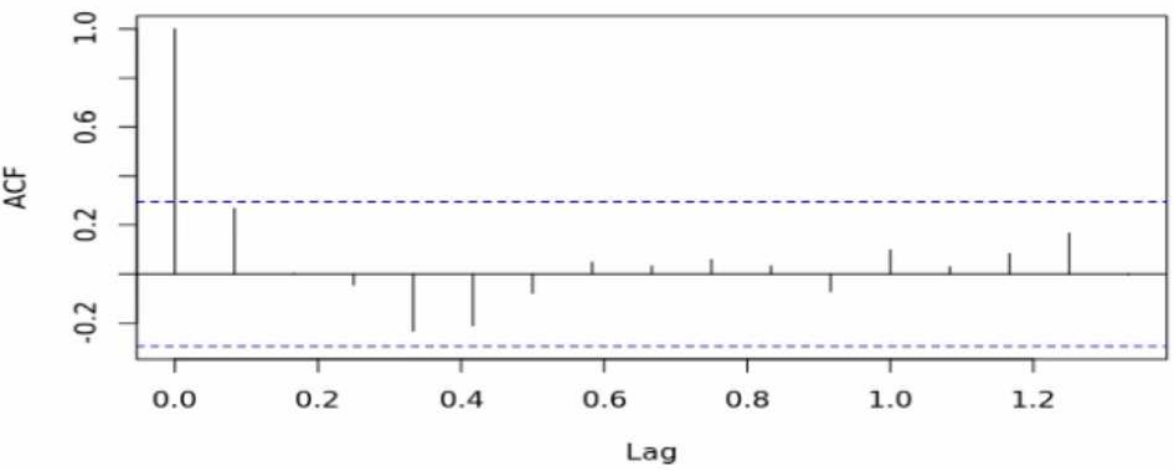

Figure 40 - Non-Remote Cards Issuer ACF

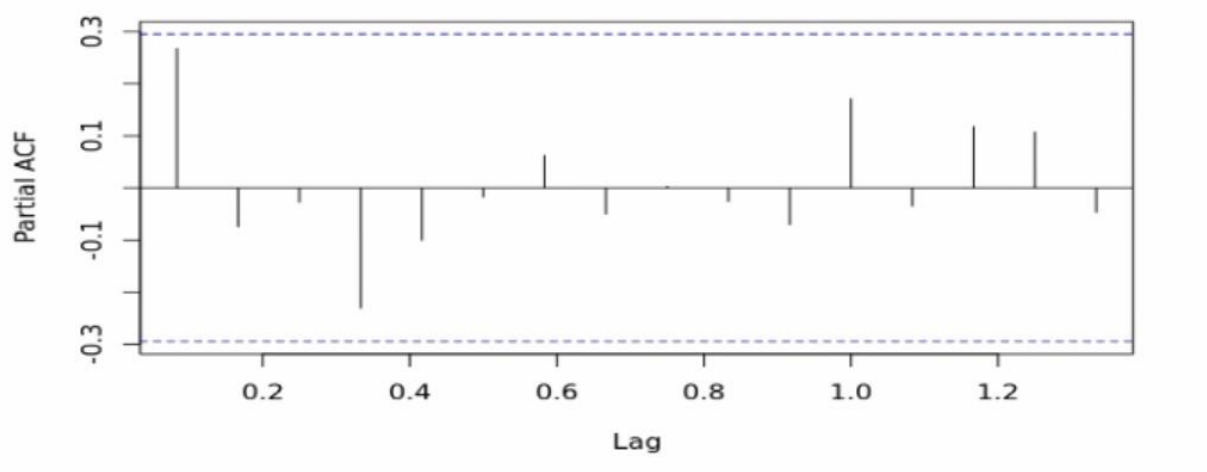

Figure 41 - Non-Remote Cards Issuer PACF

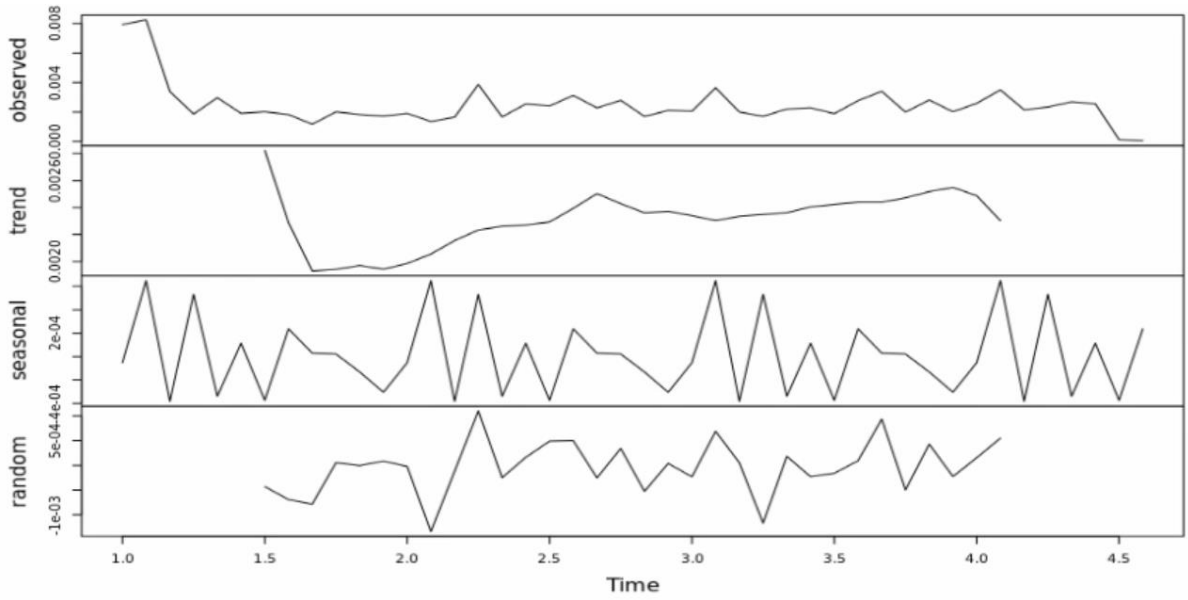

Figure 42 - Cash Withdrawals Time Decomposition

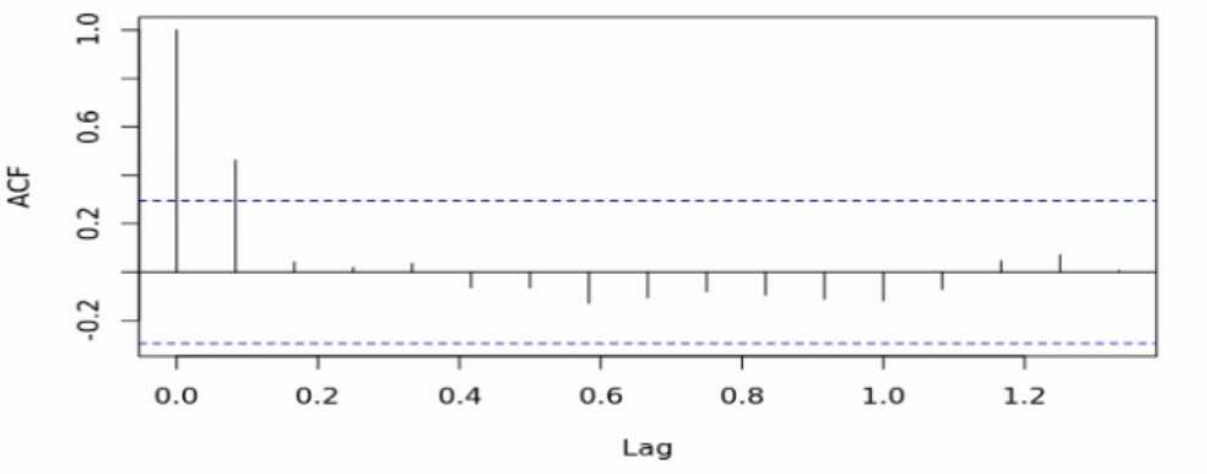

Figure 43 - Cash Withdrawals ACF

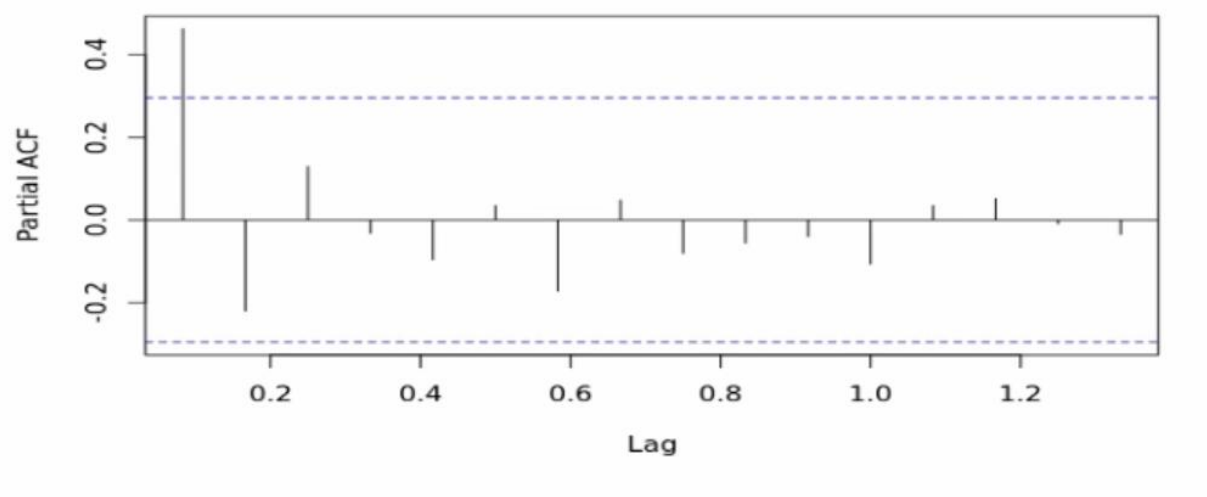

Figure 44 - Cash Withdrawals PACF