WORKING PAPERS IN RESPONSIBLE BANKING & FINANCE

Relationship Banking: The Borrower's Incentives Channel

By *Pejman Abedifar, Soroush Kamyab, Steven Ongena, and Amine Tarazi*

Abstract: We contribute to the relationship banking literature by uncovering the impact of a prior banking relationship on borrower's incentives to avoid default. As an identification strategy we exploit a proprietary dataset comprising 149,230 mortgage loans tracked monthly over a two-year period in a unique institutional setting that allows us to isolate the influence of borrower's incentives. Our findings indicate that a pre-existing relationship diminishes borrower's default risk by approximately 4%, exclusively attributable to the value of the relationship for the borrowers. This effect persists even during the notable surge in loan defaults during the COVID-19 pandemic. Our results also show that the impact of pre-existing banking relationships on avoiding default is stronger for wealthier, more religious, and male borrowers.

WP Nº 25-004

1st Quarter 2025

University of St Andrews
Scotland's first university

600 YEARS
1413 – 2013

Relationship Banking: The Borrower's Incentives Channel

Pejman Abedifar Soroush Kamyab Steven Ongena Amine Tarazi *

This draft: November 2024

A B S T R A C T

We contribute to the relationship banking literature by uncovering the impact of a prior banking relationship on borrower's incentives to avoid default. As an identification strategy we exploit a proprietary dataset comprising 149,230 mortgage loans tracked monthly over a two-year period in a unique institutional setting that allows us to isolate the influence of borrower's incentives. Our findings indicate that a pre-existing relationship diminishes borrower's default risk by approximately 4%, exclusively attributable to the value of the relationship for the borrowers. This effect persists even during the notable surge in loan defaults during the COVID-19 pandemic. Our results also show that the impact of pre-existing banking relationships on avoiding default is stronger for wealthier, more religious, and male borrowers.

Keywords: *Relationship Banking, Borrower's Incentives, Mortgage Loan, COVID-19, Default Risk.*

***JEL Classification*:** G20, G21

* Pejman Abedifar (Corresponding author) is at Tehran Institute for Advanced Studies, Khatam University, Iran, and Center for Responsible Banking and Finance, School of Management, University of St Andrews, Scotland, UK. Soroush Kamyab is at Tehran Institute for Advanced Studies, Khatam University, Iran. Steven Ongena is at University of Zürich, Swiss Finance Institute, Katholieke Universiteit Leuven, NTNU Business School, and CEPR. Amine Tarazi is at Université de Limoges, and Institut Universitaire de France, France.

We extend our gratitude to Bo Becker and the participants of the 2024 International Conference on Finance, Accounting, and Banking (ICFAB) for their valuable comments and suggestions. We also thank Javad Kashizadeh for his insightful feedback and efforts in data collection.

E-mail addresses: pa31@st-andrews.ac.uk (P. Abedifar); s.kamyab@teias.institute (S. Kamyab); steven.ongena@df.uzh.ch (S. Ongena); amine.tarazi@unilim.fr (A. Tarazi).

1. Introduction

Relationship banking and its role in default behavior are widely recognized to operate through three primary channels: screening, monitoring, and borrower's incentives (Puri et al., 2017). The motivation behind the borrower's incentives channel stems from the borrower's preference for preserving the value of the relationship they have built with their bank over time. The relationship helps clients gain easier access to bank loans, often with better terms (Peterson and Rajan, 1994; Berlin and Mester, 1999; Boot, 2000; Puri et al., 2017 among others).

While screening and monitoring mechanisms have been extensively studied (e.g., Diamond, 1984; Manove et al., 2001; Keys et al., 2010; Puri et al., 2017; Bedayo et al., 2020), the borrower's incentives channel remains relatively underexplored due to the challenges in distinguishing its impact from that of other channels. It is worth noting that some researchers have attempted to address this challenge. For instance, Schoar (2012) conducts a firm-level randomized experiment where loan officers increase personal interactions with a random group of borrowers (the treatment group). The study demonstrates that borrowers with more personal interactions with loan officers exhibit stronger incentives to avoid default.[1]

Our paper aims to address this challenge by elucidating the extent to which a prior relationship with a bank influences a borrower's incentives to avoid default. Unlike previous research, we focus on mortgage loans and differentiate between borrowers with a prior relationship with the lending bank and those without, rather than comparing the borrowers with personal interactions with loan officers to those without. This is because we believe that developing a bank-borrower relationship over time involves many factors beyond just personal interactions.,

Our study leverages a unique setting where all borrowers – with and without a prior relationship – undergo identical screening and monitoring processes, thereby enabling us to isolate the impact of borrower's incentives on loan default. This approach allows us to observe the effect of a prior relationship exclusively on a borrower's incentives in defaulting on a mortgage loan.

We exploit the distinct institutional framework in Iran's mortgage market which enables us to separate the role of borrowers' incentive from the roles of screening and monitoring. In Iran, there is only one specialized bank for mortgage loan, known as "Maskan Bank". The bank issues mortgage papers in the stock market to fund its business. According to this setting, applicants must purchase mortgage papers from the market to be eligible for mortgage loans,[2] and the bank is obliged to grant credits to the holders of these mortgage papers. In addition, all

[1] Puri et al. (2017) use the activity score – a comprehensive measure of a borrower's checking account activity – to assess a borrower's incentives to avoid default. This score includes factors such as the number of checking accounts, account balances, credit line usage, credit limit violations, and bounced checks, weighted for their predictive power of defaults. However, their study neither distinguishes borrowers with a bank relationship from those without one, nor does it fully separate the borrower's incentives channel from the screening and monitoring channels.

[2] The loan purposes encompass various activities related to housing, including purchasing, construction, renovation, and repairing a house.

mortgage-paper-holders, independent of having any prior relationships with Maskan Bank, are treated similarly in loan screening and monitoring processes. This unique setting – resembling an experiment trial – ensures that any difference in a borrower's repayment behavior is attributable solely to factors beyond monitoring and screening: the borrower's incentives channel.

In this study, we utilize a comprehensive dataset of 149,230 mortgage loan monthly performance records from March 2019 to February 2021. This dataset offers detailed insights into both loans and borrowers' characteristics, including information on pre-existing relationships between borrowers and the bank.

Our analysis shows that the borrowers with prior relationships exhibit approximately a 4% lower risk of default compared to those without such relationships. The findings underscore the importance of a borrower's incentives to avoid default and, consequently, maintain the relationship developed with the bank. Furthermore, borrowers who lack a prior relationship with Maskan Bank generally maintain relationships with other commercial banks, making default on Maskan Bank's loans less consequential for them. This strategy aligns with the literature, which suggests that borrowers may prioritize protecting relationships with their banks. As a result, they might selectively choose which obligations to fulfil and which bank to default on (Schäfer, 2019).

We also examine the relationship during the COVID-19 pandemic. The analysis reveals that there is a significantly lower default rate during the COVID-19 pandemic among borrowers with pre-existing relationships. These findings underscore the pivotal role of relationship banking and borrowers' incentives in loan performance during crises times.

Next, we examine whether the value of relationship and the incentive to maintain it varies across different clientele. Specifically, we investigate the role of wealth, religiosity, gender and marital status. On the one hand, it can be argued that banking relationship is more valuable for wealthier borrowers as they are more likely to re-use it in the future. On the other hand, these borrowers might assign less value to relationship banking given that they might have a better hard information. Therefore, we investigate whether the value relationship banking diminishes with wealth. The result of our analysis shows that wealth strengthens the relationship between prior banking relationships and loan performance suggesting that the value of banking relationship increases with wealth.

The extant literature indicates that more religious individuals are often more socially inclined (Ellison, 1992; Dunfield, 2014; Anadriani and Sabatini, 2015; Tian, 2022). This social orientation may naturally strengthen their relationships with their banks. Accordingly, maintaining their banking relationships may carry particular significance for religious borrowers. Our results indicate that the influence of prior relationships on default avoidance increases with religiosity of the borrowers.

Since women face discrimination in access to credit (Bellucci et al., 2010; Alesina et al., 2013; Ongena and Popov, 2016), prior relationships can provide them greater access to loans compared to men (Abedifar et al., 2024). Consequently, one might expect banking relationships to be more valuable for female borrowers than male borrowers. However, given that men are often more likely to engage in future business dealings, they might also benefit significantly from strong banking relationships. To investigate this, we examine the intermediary effect of gender on the relationship between prior banking relationships and loan performance. Our findings indicate that this relationship is stronger for male borrowers relative to female borrowers, suggesting that banking relationships might offer more potential benefits for men.

The potential benefits of banking relationships might depend on clients' marital status. It can be argued that banking relationships is more valuable for married borrowers because the benefits of relationships could potentially extend to their spouses and dependents. This extension could encourage them to fulfill their financial commitments better than single borrowers. Our investigation, however, provides no evidence to support this claim.

This study contributes to the existing literature on relationship banking in several ways. Our unique institutional setting enables us to investigate the role of the borrower's incentives in loan repayment using a sample of borrowers who are subject to similar screening and monitoring processes. Our analysis provides strong evidence in support of the incentive channel brought up by Puri et al. (2017). We also extend Schoar's (2012) research on the role of personal interaction of loan officers in loan performance by emphasizing the importance of prior banking relationship.

This paper also contributes to the ongoing policy debate on the role of branch banking in the economy (Burgess and Pande, 2005; Agarwal and Hauswald, 2010; Bruhn and Love, 2014; Nguyen, 2019; Levine et al., 2020). While banks have been closing branches at an unprecedented pace, extensive research demonstrates the significant influence branches have on borrowers' performance and behavior (Garmaise and Moskowitz, 2006; Bonfim et al., 2023; Benincasa, 2024). Our study further provides robust evidence to support this perspective. Additionally, we shed light on the role of banking relationship during the COVID-19 pandemic, which offers new insights into the dynamics of borrowers' behavior during an economic turmoil. Finally, we demonstrate that the value of banking relationship varies across different customer segments. In particular, our findings suggest that banking relationship is more valuable for wealthier, more religious and male borrowers.

The results of this study suggest that banks can increase borrowers' incentives to avoid default by enhancing the value of relationship banking. This can be achieved by improving the reusability of relationship banking and emphasizing its long-term benefits. For instance, banks can reward borrowers with strong repayment histories by offering better loan terms, tiered loyalty benefits, and priority access to other financial services. These policies provide borrowers with greater incentives to maintain a positive track record with the banks. This aligns with Padilla and Pagano (1997), who argue that banks can encourage borrowers to exert sufficient effort to properly implement their projects by committing to share credit information.

The remainder of the paper is structured as follows: Section 2 outlines the unique institutional setting of our study. In Section 3, we detail the methodology employed to assess the impact of the relationship on borrowers' incentives. Section 4 elucidates our data collection process and summary statistics. The findings of our analysis are presented in Section 5. To ensure the robustness of our findings, Section 6 presents a series of robustness checks. Section 7 provides further analysis of our study. Finally, Section 8 offers concluding remarks.

2. Institutional Setting

In this study, we exploit a very unique mortgage market setting under a specific regulation in Iran. According to this regulation, the housing bank issues mortgage papers to the market and individuals seeking mortgage loans must first acquire these papers from the market. Additionally, the bank is required to provide loans solely to individuals holding these mortgage papers, regardless of any prior relationship with them.

In Iran, the singular banking institution designated for housing finance is "Maskan Bank".[3] This bank provides mortgage loans under two distinct programs.[4] The maximum value of mortgage loans under either program is equivalent to US$12,000.[5] Repayment terms for both loan types span 12 years, with monthly installment payments. The programs are as follows:

First Program: Maskan Bank periodically issues a predetermined quantity of mortgage papers, typically ranging between 1 and 4 million units. These papers mature two years after issuance, and are traded in the market. Individuals seeking loans are required to procure a certain number of papers via the limit order book system and subsequently submit them to Maskan Bank's branches. Upon acquisition, individuals have the option to either utilize the papers to obtain a loan or sell them on the securities market.[6]

As an example, consider that for every 5,000,000 Iranian Rials (equivalent to $25 during the study period), borrowers are required to acquire one paper from the market, with the market price sets at $5 per paper. Therefore, for a loan amounting to the equivalent of $10,000, applicants would need to purchase 400 papers from the market, incurring a total cost of $2,000. It is noteworthy that the market price of these mortgage papers fluctuates, with the average price standing at approximately $2.76, and a standard deviation of $0.78 throughout our investigation period (see Figure 1). The highest and lowest prices recorded are $4.31 and $1.94,

[3] Besides mortgage lending, Maskan Bank provides usual financial services such as deposit account services similar to commercial banks.

[4] It is worth noting that these loans are not securitized.

[5] During the study period, US$1 is – on average – equivalent to 200,000 Iranian Rials.

[6] It is notable that each client can get mortgage loan only once.

respectively. Consequently, while the official interest rate for mortgage loans typically remains at 17.5%, the effective rate fluctuates based on market demand for papers, ranging from 17.92% to 22.97%, with a historical mean of 20.41%.[7] While the owners of mortgage papers retain the option to utilize them to secure loans or sell them in the market, buyers of mortgage papers are not allowed to resell them for a specified period, typically four months. This restriction serves to prevent market speculation and ensures that the prices of these papers primarily reflect the demand for housing loans. Figure 1 illustrates the monthly average price trends of these papers throughout the study period.

[Insert Figure 1 About Here]

The key feature of this regulation is that, regardless of clients' credit risk history, the bank is compelled to provide mortgage loans to them if they acquire mortgage papers. This regulation ensures uniform lending practices by standardizing loan terms for all borrowers. Every borrower pays the same nominal interest rate (17.5%) over the same repayment period (144 months), regardless of their credit risk. Furthermore, all borrowers – irrespective of whether they have any prior relationship with the bank – are subject to the same monitoring procedures. Upon reaching a 90-day delinquency period, the bank is mandated to reach out to the clients.[8] Subsequently, if payments remain overdue for 180 days, the bank commences the legal process to seize the collateral.[9] By implementing uniform screening and monitoring processes across all customers, we can effectively isolate the impact of prior relationships with the bank on clients' repayment behavior.

Second Program: Prospective borrowers are required to deposit half of the loan amount into their bank accounts for a duration of one year, without accruing any interest, as a prerequisite for obtaining the loan. Notably, the official loan rate under this scheme stands at 8%.[10] In this study, we will focus exclusively on the first program to investigate the borrowers' incentive channel.

It is worth noting that the repayment installments for both programs place considerable strain on Iranian households' financial capabilities, given their income levels. Figure 2 below

[7] It is worth noting that Iran's average inflation rate over the study period was approximately 30%.

[8] Based on our conversation with a loan officer at one of Maskan Bank's branches, we can confirm that the bank does not differentiate its treatment of borrowers based on prior relationships. For example, the bank does not deduct installments from alternative accounts of borrowers with pre-existing relationships, nor does it give them advance notice of potential defaults. This contrasts with the previous literature, which suggests that relationship banks tend to offer financial flexibility to their borrowers, aiming to secure future business and maintain exclusive relationships (Degryse, Ioannidou and von Schedvin, 2016; Schäfer, 2019).

[9] Borrowers' deposits in the bank are protected and cannot be used to settle the debt, thus the borrowers' account balance does not affect the borrowers' decision to default.

[10] Taking into account the 30% inflation rate during the study period, the effective loan rate escalates to 12%.

illustrates the installment-to-income ratio for Iranian households across diverse income brackets.

[Insert Figure 2 About Here]

3. Methodology

In this section, we delineate the methodology used to investigate the impact of borrowers' incentives on the probability of default based on a sample of borrowers with a prior relationship with the bank vis-à-vis those without a prior relationship. We estimate the following specification using the Linear Probability Model (LPM), probit, and logit:

$$P(Default)_i = \alpha_i + \beta_1 \, RelationshipDummy_i + \beta_2 \, Characteristics_i + \beta_3 \, RiskMeasures_i + \epsilon_i \qquad (1)$$

where *i* represents the borrower.

In this model, the dependent variable is the borrower's default occurrence within the full observation period[11] after loan origination where default is defined as failure in repaying installments for three consecutive months. If the borrower records a default during the study period, $Default_i$ takes the value of one, and zero otherwise. On the right-hand side, we categorize independent variables into three distinct sets. The variable of our interest is the prior relationship ($RelationsipDummy_i$), defined as a dummy variable that takes the value of one if the borrower *i* holds a savings or current account before obtaining the loan, and zero otherwise. The marginal effect associated with this variable provides a measure of the comparative probability of default between borrowers with a prior relationship and those without such a relationship. We control for borrower characteristics ($Characteristics_i$), including gender, marital status, occupation, and educational background. Lastly, we include variables in our model variables that indicate borrower and loan risks ($RiskMeasures_i$), such as loan-to-value ratio (*Loan to Value*), historical defaults in previous loans by borrowers (*Default in Previous Loans*), records of bounced checks (*Bounced Check*), and mortgage effective rate (*Effective Rate*). The standard errors are clustered at city level and are adjusted for heteroscedasticity and serial correlations.

4. Data and Summary Statistics

We collect proprietary data on mortgage loans, as well as borrowers' characteristics and account features, from Maskan Bank, and compile two datasets.

[11] We also use different default horizons of 6-month and 12-month for the dependent variable as a robustness check in subsection 6.3.1.

The mortgage loan dataset is organized as a panel, tracking monthly repayment status of loan contracts issued between March 2019 and February 2021. This dataset includes key loan characteristics such as the mortgage origination date, the price of mortgage paper purchased by the borrower, loan amount, collateral value, number of paid installments, and details of the branch such as its name, code, and location (city) of the branch. Since borrowers acquire mortgage papers at different prices (as illustrated in Figure 1), we calculate the effective rate for each loan based on the loan interest rate (17.5%) and the mortgage paper price.

The second dataset comprises borrowers' characteristics and account features. Borrowers' characteristics encompass a variety of attributes, such as gender, marital status, occupation, and education level.[12] In addition to personal information, risk measures are also included in the borrowers' characteristics dataset, which consist of whether the borrower has a recorded default on previous loans or a bounced check record at any bank in Iran. Account features include the type of account,[13] monthly balance, and the dates of account opening and closing. The account opening date allows us to determine whether the borrower has a prior relationship with the bank. Specifically, the relationship indicator is set to one if the borrower has an existing savings or current account prior to applying for the loan, and zero otherwise.

We merge the borrowers' characteristics and account features dataset and the mortgage loan dataset, based on the hash code of borrowers' National IDs presented in both datasets. The final dataset comprises loans and borrowers' characteristics, and performance variables for approximately 180,000 mortgage loans initiated under the first mortgage lending program of Maskan Bank during the period from March 2019 to February 2021, with the first loan installment status observable in the data in April 2019.

Since we are interested in the effect of a prior relationship on the risk of default, we exclude loans originated during the last two months of the dataset, as there is insufficient time to capture borrowers' defaults. Furthermore, we drop mortgages with inadequate data, e.g., observations with records for only one or two months are excluded. The final sample comprises 149,230 loans. Afterwards, we treat outliers by winsorizing (Chambers et al., 2000) the distribution at the 1% and 99% levels, which trims extreme values on both tails.

4.1. Summary Statistics

The summary statistics of our final dataset are presented in Table 1. For a detailed description of the variables, please refer to Appendix A1.

---

[12] It is important to note that the occupation and education level variables are represented by numerical codes, where each code corresponds to an occupational group or a specific level of education. However, the exact definitions or descriptions of these numerical codes are not provided in the dataset. Thus, we could only control for occupation and education level variables by including their fixed-effects.

[13] Account types in our dataset are classified into two main categories: savings and current accounts.

[Insert Table 1 About Here]

As illustrated in Table 1, the variable related to prior relationships (*RelationshipDummy*) indicates that 65% of loan applicants do not have a savings or current account with Maskan Bank before applying for a mortgage loan. The variable representing relationship duration (*RelationshipDuration*) shows that the average duration of a relationship that a borrower has with Maskan Bank is 9.33 years among borrowers with prior relationships. Moreover, the default indicators reveal that around 25% (and 5%) of borrowers experienced at least one recorded default defined as a 3-Month Default (and 6-Month Default), respectively.[14] Notably, the remaining borrowers exited the dataset without exhibiting any defaults.[15]

The loan to value ratio (*Loan to Value*) is, on average, 34.56%, ranging between 4.13% and 83.33%. The effective rate of loans (*Effective Rate*) averages 20.42% spanning from 17.91% to 22.97%, reflecting changes in demand for mortgage papers.

Examining the risk indicators in our data, namely the recorded bounced checks of borrowers (*Bounced Check*) and previous loan default records (*Default in Previous Loans*), we find that approximately 3% of borrowers have at least one bounced check and 12% have at least one recorded default in their previous loan contracts. Lastly, the data shows that approximately 58% of individuals in our dataset are male (*Male*), and 86% of them are married (*Married*).

4.2. Covariate Balance

Ensuring well-balanced covariates between borrowers with and without prior relationships is crucial to guarantee that the two groups of borrowers might also be similar in unobservables, and therefore, the results are not spurious. Adopting the methodology suggested by Imbens and Rubin (2015), we utilize three metrics to assess the overlap in covariate distributions:

1) **Normalized Difference:** This measure is defined as the difference in means between the two groups, divided by the square root of the average of the variances of the two samples. Mathematically, it is expressed as: $\Delta_{nr,r} = (\mu_r - \mu_{nr})/(\sqrt{(\sigma_r^2 + \sigma_{nr}^2)/2})$, where μ, σ, r, and nr denote mean, standard deviation, the sample of borrowers with prior relationship, and the sample of borrowers without prior relationship, respectively.

[14] We define "6-Month Default" as a borrower's failure to repay installments for 6 consecutive months. However, our main dependent variable in our analysis is "3-Month Default", while we use "6-Month Default" as a robustness test in subsection 6.3.3.

[15] However, it is plausible that defaults occurred post-February 2021. These are known as "right-censored subjects". We will address right-censored subjects in subsection 6.1.1 using the Cox model by formulating the log-likelihood function as a weighted combination of two components: the sample density of loans experiencing default within the study time and the survivor function representing the censored loans.

2) **Dispersion Difference:** This measure is defined as the logarithm of the ratio of the standard deviations of the sample of the borrowers with prior relationships to the sample of the borrowers without prior relationships. It is mathematically represented as: $\Gamma_{nr,r} = \log{(\sigma_r/\sigma_{nr})}$.

3) **Coverage Frequency:** This measure evaluates the degree of overlap between the two groups by assessing the proportion of borrowers without prior relationships (borrowers with prior relationships) units that have covariate values in the tails of the distribution of that covariate in the borrowers with prior relationship (borrowers without prior relationship) units. It is represented as: $\pi_r^\alpha = (1 - F_r(F_{nr}^{-1}(1-\alpha/2))) + F_r(F_{nr}^{-1}(\alpha/2))$, where for the specific value of α, π_r^α denotes the probability mass of the covariate's distribution in the sample of the borrowers with prior relationships that is located outside the quantiles $1-\alpha/2$ and $\alpha/2$ of that covariate in the sample of the borrowers without prior relationships. This measure ensures that the tails of the covariate distribution in one group do not extend beyond the extreme values observed in the opposite group, and vice versa.

Additionally, to evaluate the overall balance between the two samples, we estimate the propensity score (*P-Score*) and its log odds ratio (*Linearized P-Score*) using a logistic regression where the relationship indicator (*RelationshipDummy*) is regressed on both continuous and categorical covariates. By calculating the overlap measures, i.e. the normalized difference, dispersion difference, and coverage frequency for the estimated propensity scores, we can assess the balance between the two groups of borrowers.

[Insert Table 2 About Here]

As demonstrated in Table 2, the distributions of all variables are nearly identical between the two samples. The statistics for the normalized difference and dispersion difference for all variables are sufficiently close to zero. Additionally, the coverage frequency at the 95% level (α = %5) is also close to zero for all variables. The normalized difference, dispersion difference, and coverage frequency measures for P-Score in sample of the borrowers without prior relationship (with prior relationship) are 0.24, 0.10, and 0.05 (0.06), respectively. These statistics for the linearized P-score are 0.24, 0.10, and 0.05 (0.06).

The balanced distribution of both individual covariates and the estimated propensity score (including the log odds ratio of the propensity score) ensures that the two sub-sample of borrowers are (hopefully) similar in unobservable factors, and our inference is not based on extrapolation due to a lack of corresponding borrowers in the other sample.

5. Results

In this section, we discuss the results of estimation of Equation (1) using LPM, probit, and logit estimation techniques as our statistical methods for exploring whether there is any linkage between prior banking relationships and borrower's incentives in payment of loan installments. The average marginal effects are presented in columns (1), (2), and (3) of Table 3, respectively.

[Insert Table 3 About Here]

Our main variable of interest, the *RelationshipDummy*, exhibits a negative effect across all three models, indicating that borrowers with pre-existing relationships with the bank are less likely to default on their loans compared to those without such relationships. Specifically, the average marginal effect of prior relationships, as reported in columns (1) to (3), is -0.0419, -0.0423, and -0.0429, respectively, all of which are statistically significant at the 1% level. This indicates that having a prior relationship lowers default risk by about 4.2%, exclusively attributable to the borrower's incentives. The impact is also economically meaningful given that the average default rate is 34.56%. The findings on borrowers' characteristics – gender (*Male*) and marital status (*Married*) – exhibit insignificant coefficients across all three models.

The results consistently indicate a significant positive association between the borrower's risk measures – *Loan to Value*, *Bounced Check*, and *Default in Previous Loans* – and the probability of default across all models. Specifically, a one standard deviation increase in the *Loan to Value* ratio increases the probability of default by approximately 1%. The presence of at least one bounced check in borrower's record history is associated with a substantial increase in default risk, raising it by 15.90%, 14.03%, and 14.13% in columns (1), (2), and (3), respectively. The average marginal effect of having a default history in previous loans (*Default in Previous Loans*) is shown to significantly increase the likelihood of default by 50.89%, 39.41%, and 37.65% in columns (1), (2), and (3), respectively. The negative coefficient on the *Effective Rate* suggests that borrowers who purchase mortgage papers at higher prices exhibit a lower inclination to default, which aligns with expectations, as these borrowers likely have a stronger motivation to secure their mortgage.

Comparing the average marginal effect of the *RelationshipDummy* with other variables reveals the substantial economic significance of having prior relationships. For instance, as shown in Table 3, the approximate 4% impact of prior relationships on default risk is notably more than the impact of *Loan to Value*. Specifically, one standard deviation increase in the *Loan to Value* only raises the default risk by 1%, highlighting the greater economic significance of maintaining prior relationships with the bank. However, the borrower's history, particularly records of bounced checks (*Bounced Check*) and previous loan defaults (*Default in Previous Loans*), emerges as more economically significant, increasing the probability of default by an average of 14.68% and 42.65%, respectively, across all models.

In summary, the findings derived from the estimations across the three columns consistently highlight that prior relationships between the borrower and the lender significantly reduces the borrower's inclination to partake in moral hazard, thereby decreasing the likelihood of defaulting on the loan. These outcomes emphasize the protective function of pre-existing relationships with the lending institution in reducing default occurrences, underscoring the value of relationship banking for borrowers.

Our findings align with the research conducted by Schoar (2012), which shows that borrowers who receive heightened personalized engagement with the bank demonstrate significantly better loan repayment, characterized by decreased delinquency rates, fewer delinquency occurrences, and a delayed onset of initial default. Furthermore, Puri et al. (2017) exhibit a reduction in borrower's incentives to default due to the perceived value of the relationship, a finding consistent with our results. The results also align with existing literature (Schäfer, 2019), which indicates that borrowers strategically choose to default on banks that pose fewer consequences for them.

6. Robustness Checks

In this section, we conduct a battery of robustness tests. First, we use alternative estimation methods including Cox proportional hazard model as a semi-parametric survival analysis and machine learning estimation methods such as random forest, XGBoost, and neural network. Second, we address endogeneity concerns in three ways. We explore whether establishing relationship with Bank Maskan is mainly derived by geographical distance. This provides suggestive evidence that the assignment of relationship with the bank is exogenous to the individual characteristics. Next, we control for borrowers' wealth, by using a sub-sample of borrowers domiciled in Tehran, to make sure that our results are not driven by differences in wealth. Last, we address the endogeneity concern by limiting the sample of individuals to those who already have relationships with Maskan Bank. This approach helps us exclude the possibility that our results are driven by unobservable differences between borrowers with prior relationships and those without. Third, we examine the robustness of our results using alternative default measures.

6.1. Alternative Estimation Methods

6.1.1. Survival Analysis

In our baseline investigation, we employ LPM, probit, and logit regressions with the full observation period default horizon as the dependent variable. In this subsection, we use a different approach that seeks to examine how long it will take for a default to occur during the study period. We primarily employ the Cox proportional hazard model (Cox, 1972) for our analysis in this subsection to tackle and overcome the right-censorship in data.

Cox proportional hazard regression is particularly suited for analyzing time-to-event data. It is a widely adopted technique to explore the hazard rates of loan default among borrowers (e.g. Ongena and Smith, 2001; Baele et al., 2014; Agarwal et al., 2018). By estimating hazard ratios, we can assess the impact of various factors on default risk, providing valuable insights into the determinants of loan default behavior among borrowers. We follow the extant literature, and adopt the following specification:

$$h(t|x) = h_0(t).\exp\left(\beta_1\ RelationshipDummy_i + \beta_2\ Characteristics_i + \beta_3\ RiskMeasures_i\right) \quad (2)$$

where *i* represents the borrower.

This semi-parametric model allows for a non-parametric baseline hazard rate as well as potentially time-varying explanatory variables. In this model, the dependent variable is time-to-default where default is defined as failure in repaying installments for three consecutive months. Similar to Equation (1), our main variable of interest is the $RelationshipDummy_i$. The hazard ratio associated with this variable provides a measure of the comparative default risk between borrowers with a prior relationship and those without such a relationship. The control variables are the same as those used in Equation (1).

Full repayment of a loan or the conclusion of the sample period can result in the absence of default observation for that particular loan. In the latter case, such loans are considered "right-censored", as whether the borrowers ultimately default on them remains unknown. In the presence of right-censoring, accurately estimating the time to default becomes problematic and failure of controlling for right censoring can result in biased and inconsistent estimates of the model parameters. To tackle this issue, right-censored observations could be incorporated in the estimation by formulating the log-likelihood function as a weighted combination of the sample density of loans experiencing defaults in the study period, and the survivor function representing the right-censored loans. This approach ensures that both defaulted and censored loans contribute to the estimation process, yielding consistent parameter estimations.

Based on the information about the loan's initial installment date and the monthly loan performance, we construct a survival data where the loan contract date marks the initial time point and the occurrence of the first loan default represents the event of interest. Each individual (borrower) is enrolled and followed for a specified duration, denoted as the subject's follow-up period. According to our dataset, the shortest follow-up period until the first default is observed at three months, whereas the longest duration until a borrower defaults for the first time is 23 months post-loan acquisition (see Table 4). As indicated in Table 4, a total of 109,706 borrowers had no default incidents during their follow-up period, with an average duration of 11.6 months. However, 39,524 borrowers had at least one default incident during their follow-up period, with an average duration of 6.75 months.

[Insert Table 4 About Here]

To gain a preliminary understanding of the dataset, Figure 3 depicts the time-to-event follow-up periods for a random sample of 50 borrowers, indicating whether they default or depart the study window with a clean payment record. As we can see in Figure 3, all subjects remain in the dataset until the conclusion of the period unless a default event is recorded. This suggests that the sole reason for right censoring in our dataset is the conclusion of the study period. It is also important to highlight that our survival data exclusively demonstrates right censoring, with no instances of left or interval censoring, as we consistently observe loans originated after March 2019 on a monthly basis without any missing data points during the borrowers' follow-up period. Additionally, there are no truncations in our data. It is important to note that while this figure showcases only a small sample of our data, the aforementioned characteristics hold for the entire dataset.

[Insert Figure 3 About Here]

An intriguing aspect of the data is the duration until the first default. The data reveals that the highest number of defaults happens between 3 and 4 months after the borrower acquires the loan, afterwards, the frequency of first defaults decreases as time passes. Figure 4 presents a histogram for the two groups of borrowers, one who have a default during the study time and one who leave the study period without any defaults. It is evident that there are relatively few borrowers who manage to avoid default as time elapses. The majority of these borrowers remain in the dataset for 11 to 13 months before they exit the study window.

[Insert Figure 4 About Here]

Table 5 presents the results of our analysis. In column (1), we focus solely on the variable of interest, the *RelationshipDummy*, which is included as the explanatory variable. In column (2), we expand the model by introducing control variables that capture borrowers' characteristics, such as gender (*Male*) and marital status (*Married*), as well as fixed effects for job and education. Finally, in column (3), we further enhance the analysis by including additional variables that control for borrowers' risk measures, specifically *Loan to Value*, *Bounced Check, Default in Previous Loans*, and *Effective Rate*.

The hazard ratios for *RelationshipDummy* are 0.818, 0.844, and 0.804, as reported in columns (1), (2), and (3) of Table 5, respectively, all of which are statistically significant at the 1% level. These ratios indicate that borrowers who have a savings or current account before the mortgage origination date have hazard rates that are 18.2% (= 1 - 0.818), 15.6% (= 1 - 0.844), and 19.6% (= 1 - 0.804) lower than those of borrowers without such a relationship. These findings re-affirm the robustness of our previous results from LPM, probit, and logit regressions, highlighting that maintaining a prior relationship with the bank serves as a protective factor against default risk.

[Insert Table 5 About Here]

The findings on the control variables consistently reveal a positive association between default and factors such as *Loan to Value*, *Bounced Check*, and *Default in Previous Loans* across all models. In contrast, the hazard ratio of *Effective Rate* exhibits a decreasing effect on default risk. These results are qualitatively consistent with those observed in the main model, reaffirming the significance of these risk measures. On the other hand, control variables like *Male* and *Married* exhibit insignificant coefficients in column (3), aligning with our main findings.

While the results of Cox model present promising findings, it is imperative to verify whether the hazard ratios maintain proportionality over time. This assessment is critical to bolster our confidence in the validity and reliability of our conclusions. Thus, we investigate whether the proportional hazards assumption holds (Hosmer et al., 2008). This assumption posits that the hazard ratios for the covariates remain constant over time, while the hazard rates are allowed to vary over time. Violation of this assumption in Cox regression analysis can result in biased estimates (Kuitunen et al., 2021). Our tests confirm that the proportional hazards assumption holds.[16] The results of our tests are reported in Appendix A2.

Overall, the consistent findings of the Cox model provide robust evidence of the relationship between borrower's prior relationships and default behavior, enhancing the validity and reliability of our previous results.

6.1.2. Machine Learning Methods

In this subsection, we extend our analysis by applying machine learning techniques to estimate the primary model specified in Equation (1). We utilize machine learning models – specifically, random forest (Breiman, 2001; Khandani et al., 2010), XGBoost (Chen and Guestrin, 2016), and neural networks (Ciampi et al., 2013) – which have demonstrated success and popularity in predictive tasks. Following the methodology of Fuster et al. (2022), we employ these models to assess their predictive power and robustness in estimating default risk compared to the logit model.[17]

We begin by estimating all four models – logit, random forest, XGBoost, and neural network – using a portion of our full dataset, designated as the training set. To evaluate the models' performance, we use a separate test set. Specifically, 70% of the dataset is randomly selected for training, while the remaining 30% is reserved for testing. This random sampling is performed across all loans, ensuring that the division between training and test sets is independent of all characteristics.

Within the training set, we further split the data into two subsets. The first subset, representing 70% of the training data, is used to estimate all models and train the machine learning models.

---

[16] We use the tests such as: comparison of observed and predicted survival curves (Cleves et al., 2008), checking parallelness of log-log survival curves, checking parallelness of adjusted log-log survival curves, and graphical test of scaled Schoenfeld residual (Schoenfeld, 1982; Grambsch and Therneau, 1994).

[17] For a comprehensive discussion on these models, please refer to the studies by Khandani et al. (2010), Breiman (2001), Fuster et al. (2022), and Ciampi et al. (2013).

The second subset, which comprises the remaining 30% of the training data and is referred to as the calibration sample, is utilized to apply isotonic regression. This step refines the predicted class labels from the machine learning models into calibrated probabilities. By employing this approach, we ensure that both the traditional and machine learning models are based on an equivalent amount of data for estimating default probabilities. Appendix A3 presents our evaluations of the models' performance.

Unlike the logit model, which allows for straightforward interpretation of marginal effects due to its parametric nature, machine learning models such as random forest, XGBoost, and neural networks are inherently non-parametric. This non-parametric nature introduces complexity in understanding direct marginal effects, as these models capture intricate, non-linear interactions between features. Consequently, while interpreting direct marginal effects is challenging, we analyze the change in the probability of default by evaluating the impact of switching dummy variables from zero to one and assessing the effect of a one-unit change in continuous variables from their mean values. This approach enables us to compare the results of machine learning models with those from our main logit regression.

Table 6 presents the impact of various features on the probability of default. Notably, the marginal effects of *RelationshipDummy* are -3.23%, -3.95%, and -3.80% for random forest, XGBoost, and neural network models, respectively. These effects are consistent with our primary findings, which underscores the robustness and reliability of our results. This consistency generally extends to other features as well, with the exception of *Loan to Value*. While the neural network model's results align with our main logit model, both random forest and XGBoost models exhibit a negative effect for *Loan to Value*, contrasting with the positive effect observed in logit model.

[Insert Table 6 About Here]

6.2. Endogeneity Concerns

6.2.1. Random Assignment of Relationship with Maskan Bank

To ensure that the borrowers with prior relationships with Maskan Bank are not systematically different from those without relationships, we explore whether the assignment of the relationships can be considered almost random. Specifically, we test whether borrowers' proximity to Maskan Bank branches influences their likelihood of having prior relationships. If the distance to the bank is a significant factor in determining a borrower's relationship, this would support the notion that the relationship assignment is exogenous to the borrower's characteristics. In this analysis, we examine the geographical distance between a borrower's residence and the nearest Maskan Bank branch and its impact on the likelihood of having a pre-existing relationship. The idea is that borrowers living closer to Maskan Bank branches are more likely to engage with the bank simply due to convenience, rather than any inherent characteristics or selection biases.[18] This approach allows us to demonstrate that the prior relationships are likely the result of geographical factors, not selective treatment, thus supporting the almost random assignment of relationships.

We follow two approaches to analyze the link between having banking relationships and geographical distance between borrowers and Maskan Bank branches: (*i*) we count the number of Maskan Bank branches within circles of varying radii centered on the borrowers' neighborhoods, identified by their 5-digit ZIP codes, and (*ii*) we calculate the average distance to the *k* nearest branches. We use Google Maps to identify the geographic coordinates of each bank branch, as listed on the banks' websites. In Tehran, there are 32 operating banks, each with branches distributed throughout the city. The total number of bank branches in Tehran is 1,686, with 84 branches of Maskan Bank. Figure 5 illustrates the distribution of Maskan Bank branches in Tehran.

[Insert Figure 5 About Here]

Table 7 presents the distribution of bank branches (Maskan Bank and other banks) within varying radii (500, 1000, 1500, and 2000 meters) centered on borrowers' neighborhoods center. On average, borrowers have 0.20, 0.71, 1.41, and 2.38 Maskan Bank branches within 500, 1000, 1500, and 2000 meters, respectively. In contrast, there are significantly more branches of other banks, with an average of 2.97 branches within 500 meters, increasing to 41.56 branches within 2000 meters. This disparity highlights the relatively lower density of Maskan Bank branches compared to other banks in the borrowers' vicinity.

[Insert Table 7 About Here]

We first examine the effect of having additional Maskan Bank branches in the vicinity of a borrower's residence on the likelihood of having prior relationships with Maskan Bank. To do this, we define a variable that represents the number of branches within a circle centered on the borrower's 5-digit ZIP code center, with radii of 500, 1000, 1500, and 2000 meters. We then regress the *RelationshipDummy* on the number of nearby Maskan branches (NMB_i) and other borrower characteristics. These characteristics include gender, marital status, record of default in previous loans and bounced checks, job, and education, using the following logit regression model:[19]

[18] We acknowledge that neither borrower nor bank branch locations are entirely random. However, most Maskan Bank branches have been in the same location for an extended period, making it unlikely that borrowers choose their residence solely based on proximity to Maskan Bank branches. Therefore, within a certain radius, the distance between a branch and a borrower can be reasonably considered fairly random.

[19] For the sake of simplicity, we only report the results of logit regression. We confirm that we find similar results when we use LPM and probit regressions in the present subsection.

$$\mathrm{P}\left(RelationshipDummy_{\mathrm{i}}\right) = \beta_0 + \beta_1.NMB_i + \beta_2.Characteristics_i + \epsilon_i \tag{3}$$

The results presented in Table 8 reveal a statistically significant positive relationship between the number of nearby Maskan Bank branches and the likelihood of a borrower having a prior relationship with the bank. In all four models (using the four different radius circles), the coefficient on the number of Maskan Bank branches is positive and significant at the 1% level. Specifically, for the 500m circle, the presence of an additional branch increases the likelihood of a prior relationship by 2.15%. As the radius expands to 1000m, 1500m, and 2000m, the effect slightly decreases, with coefficients of 0.0131, 0.0136, and 0.0127, respectively. This suggests that the proximity of Maskan Bank branches is a crucial factor in the formation of prior relationships with borrowers, emphasizing that geographic distance plays a significant role in relationship assignment.

[Insert Table 8 About Here]

Instead of using the number of nearby Maskan Bank branches, we also examine the effect by employing another variable (MBD_i), which determines the average distance to the k nearest branches, where k belongs to the set $\{1, 2, 3, \ldots, 10\}$. We define MBD_i as:

$$MBD_i \equiv \frac{\sum_{j=1}^{k}(Distance\ to\ the\ j_{th}\ Nearest\ Maskan\ Branch)_i}{k} \tag{4}$$

Table 9 provides summary statistics for the distance between borrowers' neighborhoods and the nearest bank branches, including Maskan Bank and other banks. On average, the distance to the nearest Maskan Bank branch is 1,287.32 meters, while the average distance to the 10 nearest branches increases to 3,519.98 meters. The results indicate that Maskan Bank branches tend to be located farther from borrowers' neighborhoods compared to other banks, highlighting the relatively lower accessibility of Maskan Bank services.

[Insert Table 9 About Here]

We run Equation (3) again except we substitute NMB_i with MBD_i variable. This approach allows us to analyze how varying distances to the closest Maskan Bank branches impact the likelihood of borrowers establishing prior relationships with the bank, providing a more nuanced understanding of the influence of geographic proximity on relationship formation.

The results displayed in Table 10 reveal a statistically significant negative relationship between the distance to the nearest Maskan Bank branch (MBD) and the likelihood of a borrower having a prior relationship with the bank. In all columns, the coefficient for the distance to the nearest branches is negative and significant at the 1% level. The coefficients range from -0.0018 to -

0.0021 as we consider the average distance to the nearest branch up to the ten nearest branches. This indicates that as the distance to the Maskan Bank branch increases, the likelihood of establishing a prior relationship decreases.

[Insert Table 10 About Here]

In addition to our logit regression analysis, we employ the machine learning method of random forest to assess the feature importance of each variable, especially the borrowers' distance to the Maskan branches, in predicting whether a borrower has a prior relationship with Maskan Bank. By analyzing the importance scores generated by the model, we identify which factors, including the distance to Maskan branches, gender, marital status, and prior loan defaults, significantly influence the likelihood of borrowers establishing a relationship with the bank.

[Insert Figure 6 About Here]

The results from the random forest analysis, as shown in Figure 6, reveal the feature importance scores for predicting whether a borrower has a prior relationship with Maskan Bank. The most significant predictor is the *Distance to Maskan Branch* (*MBD*), with an importance score of 0.835, indicating that the distance to branch plays a crucial role in determining borrower relationships. Following *MBD*, *Job* has a much lower importance score of 0.081, suggesting that it also contributes to the prediction but to a significantly lesser extent. Other features, such as *Education* (0.057) and *Gender* (0.009), demonstrate even lower levels of importance. The variables *Marital Status* (0.006), *Default in Previous Loans* (0.005), and *Bounced Check* (0.004) have minimal impact on the model's predictions. These findings underscore the dominance of proximity to Maskan Bank branches in influencing borrowers' relationships with the bank.

The presence of other banks branches in the vicinity of borrowers' neighborhoods is also an important issue which can affect on the borrowers' decision for having relationship with Maskan Bank. To assess this case, we introduce a ratio (DR_i) defined as follows:

$$DR_i \equiv \frac{(Distance\ to\ the\ Nearest\ Maskan\ Branch)_i}{(Average\ Distance\ to\ the\ k\ Nearest\ Other\ Banks\ Branches)_i} \tag{5}$$

where $k \in \{1, 2, 3, \ldots, 10\}$. We hypothesize that an increase in the distance to Maskan Bank branches or a decrease in the distance to other banks' branches will lead to a lower probability of borrowers establishing a relationship with Maskan Bank. Consequently, we run Equation (3) again, substituting NMB_i with the DR_i ratio.

[Insert Table 11 About Here]

The results presented in Table 11 illustrate a statistically significant negative relationship between the distance ratio (*DR*) and the likelihood of borrowers having a prior relationship with Maskan Bank. The coefficient on *DR* consistently indicates a significant negative impact, with estimates ranging from -0.0021 for $k = 1$ to -0.0079 for $k = 10$. This suggests that as either the distance to the nearest Maskan branch increases or the distance to the other banks' branches decreases, the probability of establishing a prior relationship with Maskan Bank declines.

We also employ the random forest to assess the feature importance of variables, particularly focusing on the *DR*, in predicting whether a borrower has a prior relationship with Maskan Bank. The results as illustrated in Figure 7 show that *DR* is the most important predictor, with importance scores ranging from 0.8306 to 0.8415 across different models with different values of *k*. In comparison, other variables such as *Job* (0.0797 to 0.0879), education (0.0528 to 0.0586), and *Gender* (0.0082 to 0.0091) show significantly lower importance scores. *Marital Status* and other historical variables, such as *Bounced Checks* and *Defaults in Previous Loans*, exhibit even smaller importance values.

[Insert Figure 7 About Here]

6.2.2. Controlling for Borrower's Wealth

To ensure the robustness of our findings, it is essential to account for the potential influence of borrower's wealth on the link between prior relationships and default behavior. Specifically, there may be a correlation between having prior relationships with Maskan Bank and borrower's wealth, which could affect the coefficient on the *RelationshipDummy*. In this subsection, we utilize the average home price per square meter in the borrowers' neighborhoods at the time of loan origination as a proxy for wealth. The assumption is that individuals living in more expensive areas tend to be wealthier.

We use data on housing prices per square meter in Tehran, with accuracy based on the first 6-digit ZIP code, as illustrated in Figure 8.[20] This data is reported monthly by the Ministry of Roads and Urban Development.[21] Average home prices per square meter in Tehran range from $250 to $1,450 (see Appendix A4 for more details on home prices).[22] We merge this data with our main dataset using the 5-digit ZIP code[23] variable, which represents the borrowers' neighborhoods at the time of loan origination.

[20] The analysis in this subsection is limited to mortgage loan data for Tehran because home price data is only available for this city. Hence, we are unable to use home prices to control for wealth in our main analysis.

[21] https://www.mrud.ir/en/

[22] The data on the size of borrowers' homes is not available to compute the home value.

[23] Our dataset includes the 5-digit ZIP codes of the borrowers' locations. Since the home price data is at 6-digit ZIP code level, we calculate the average home prices per square meter at 5-digit ZIP code to make it compatible to our main dataset.

[Insert Figure 8 About Here]

The results for the *RelationshipDummy* variable, presented in Table 12, are consistent with our main findings. The negative coefficient across LPM, probit, and logit models remains statistically significant at the 1% level, indicating that borrowers with prior relationships are less likely to default. The magnitude of the effect is slightly increased, with marginal effects of approximately -4.4% across all models, which is only marginally different from the results in the main analysis. This consistency reinforces the robustness of the borrower's incentive channel, even when accounting for borrower's wealth.

[Insert Table 12 About Here]

6.2.3. The Sample of Borrowers with Prior Relationships

One might argue that the results are driven by the potential differences between borrowers with and without relationships. To tackle this concern, we study the sub-sample of borrowers who already have a prior relationship. Consequently, we replace our variable of interest, i.e. *RelationshipDummy*, to the duration of relationships (measured in years) commencing from the opening of an account with the bank and ending with the origination of the loan (*RelationshipDuration*). This allows us to measure the effect of an additional year of relationship on borrowers' incentives to avoid defaults merely among those with existing relationships. The results of LPM, probit, and logit regressions are presented in Table 13, columns (1) to (3), respectively. The coefficient on *RelationshipDuration* is statistically significant in all three columns. The results indicate that each additional year of relationship reduces the probability of default for about 0.22%, 0.20%, and 0.22%, respectively.

[Insert Table 13 About Here]

6.3. Alternative Default Measures

6.3.1. Alternative Default Horizons

Our primary analysis uses a binary dependent variable to indicate whether a borrower defaults within the full observation period following loan origination, with default defined as the failure to repay installments for three consecutive months. To explore the robustness of our findings, we consider alternative default horizons. Specifically, we re-define the dependent variable to capture defaults occurring within 6-month and 12-month periods post-loan origination, again defining default as a failure to make payments for three consecutive months. We re-estimate Equation (1) with the same explanatory variables, but using the alternative definitions of dependent variable to examine the impact of different time frames on our results.

[Insert Table 14 About Here]

Table 14 presents the results for default horizons of 6-month and 12-month periods. As shown in columns (1) to (6), the average marginal effects of having prior relationships are significantly negative across both horizons, suggesting that borrowers with prior relationships consistently have less incentives to default. Specifically, within the 6-month period after loan origination, borrowers with prior relationships show a reduced likelihood of default by -3.33%, -3.02%, and -2.89% in the LPM, probit, and logit models, respectively, compared to those without such a relationship. Similarly, the impact of prior relationships on default within a 12-month period is -4.88%, -3.02%, and -2.89% in the LPM, probit, and logit models, respectively. These results align closely with our findings when considering the full observation period for default, reinforcing the robustness of prior relationships as a protective factor against default.

The coefficients on the control variables are qualitatively similar to the main results except for the *Loan to Value* which exhibit a significantly negative relationship with the probability of default in columns (4), (5), and (6), where we consider 12 months period for default horizon.

6.3.2. Alternative Measure of Relationship

It can be posited that the duration of a borrower's relationship with a bank holds substantial sway over her motivation to fulfill installment payments. For example, clients with lengthy history of engagement with a bank might demonstrate heightened allegiance and wield greater bargaining power. Consequently, they may prioritize nurturing a positive rapport with the bank, thereby demonstrating increased diligence in avoiding defaulting on their obligations. To delve deeper into this notion, we undertake a robustness assessment by substituting the binary relationship indicator with a continuous variable representing the duration of the borrower's affiliation with the bank.

This analytical approach enables us to probe whether the depth or duration of the relationship exerts a noticeable impact on default tendencies, furnishing nuanced insights into the underlying intricacies of borrower-bank interactions. Consequently, we *RelationshipDuration* instead of *RelationshipDummy* in our regressions.

[Insert Table 15 About Here]

Based on the findings presented in Table 15, it is evident that each additional year of relationship with the bank corresponds to a decrease in default risk in all estimation methods. The models suggest that a one-year increase in the relationship duration decreases the risk of default by about -0.32%, -0.31%, and -0.33% in LPM, probit, and logit, respectively. These effects are larger than the results in Table 13 which is because of including the borrowers without prior relationships in this subsection.[24] This underscores the importance of the depth of the borrower-bank relationship in influencing the borrower's inclination to avoid default. The result is also consistent with Fiordelisi et al. (2014), which highlights the benefits of longer bank-borrower relationships in reducing the probability of default.

6.3.3. Weak-Default vs Hard-Default

In our initial analysis, default is defined as being delinquent for 90 days in repayment, denoted as "weak-default". This criterion entails marking the failure event when borrowers have surpassed a 90-day period without making any repayments on their loans. An alternative metric for default, termed "hard-default", stipulates default as occurring after 180 consecutive days without any repayments (Puri et al., 2017). To ensure the reliability of our results, we adjust the dependent variable to the hard-default definition and re-estimate Equation (1). This method enables us to evaluate the consistency of our findings with a different default definition and validate the robustness of our primary results. Furthermore, it allows for a deeper understanding of borrowers' behaviors concerning the two distinct types of defaults: weak-default and hard-default.

[Insert Table 16 About Here]

As shown in Table 16, borrowers with a pre-existing relationship demonstrate a significant disincentive toward hard-default. Specifically, the average marginal effect of having prior relationships (*RelationshipDummy*) suggests that the risk of default is approximately 1.64% to 1.73% lower for borrowers with relationships compared to those without. These results are qualitatively similar to our main results when we use weak-default as the dependent variable; however, it indicates that prior relationships exert a more significant economic impact on preventing weak-defaults compared to hard-defaults.

7. Further Analysis

7.1. COVID-19 Lockdown and Relationship Banking

The COVID-19 pandemic and subsequent lockdowns brought about unprecedented challenges for households worldwide, causing an exogenous income shock (Bruce et al., 2022). Such an unforeseen income shock has significant implications for borrowers' ability to meet their loan obligations. Previous studies indicate a surge in loan defaults in the immediate aftermath of the COVID-19 lockdown (Nigmonov and Shams, 2021). Our sample presents a similar trend. In

[24] Inclusion of borrowers without prior relationships in the sample magnifies the effects, as having prior relationships alone emerges as a significant factor in shaping a borrower's incentives.

Iran, the lockdown began in March 2020. Figure 9 shows the trend in the number of defaults each month, spanning 11 months prior to the lockdown period until 11 months thereafter.

[Insert Figure 9 About Here]

As we can see, following the onset of the lockdown in March 2020, there is a notable increase in defaults. This sharp rise reflects the financial strain experienced by borrowers due to pandemic-related disruptions to household income and economic activity. However, as time progresses beyond the initial shock period, there is a gradual decline in defaults, suggesting a diminishing trend in default rates. This decrease may signify a partial recovery or adaptation of borrowers to the new economic realities imposed by the pandemic.

Within the realm of banking, the role of pre-existing relationships between borrowers and financial institutions during such crises is unclear priori. Indeed, there are two competing hypotheses:

Positive Effect: Borrowers may perceive their pre-existing relationships with the bank as valuable assets during the pandemic. They may strive to preserve these relationships, anticipating potential future benefits in accessing financial support if the crisis worsens. This perspective suggests that borrowers with established relationships may be more inclined to honor their loan commitments despite the financial difficulties induced by the pandemic.

No Effect: Conversely, pre-existing relationships might not significantly influence borrowers' default decisions during the pandemic lockdown. Borrowers facing financial distress may view the crisis as a justification for defaulting, regardless of their relationship status with the bank. This perspective implies that the impact of relationship banking on loan default behavior remains negligible in the face of severe economic disruptions.

To empirically examine the influence of pre-existing relationships on loan default behavior during the COVID-19 lockdown, we conduct an event-study to incorporate the pandemic's temporal dynamics.[25] The regression model takes the following form with the same explanatory variables in Equation (1), except that we include the interaction term between $RelationshipDummy_i$ and month dummies (T_t):

$$\begin{aligned} \mathrm{P}(Default)_{it} = \beta_0 + \sum_{t=0}^{t=23} \beta_{1t} T_t + \beta_2\, RelationshipDummy_i + \sum_{t=0}^{t=23} \beta_{3t}\, RelationshipDummy_i \times T_t \\ + \sum_{t=0}^{t=23} \beta_{4t}\, Charateristics_{it} + \sum_{t=0}^{t=23} \beta_{5t}\, RiskMeasures_{it} + \epsilon_{\mathrm{it}} \end{aligned} \tag{6}$$

[25] We use a linear probability model in this subsection. Also, we estimate the model using probit and logit regressions; however, for the sake of simplicity we only report the results of LPM regression. We confirm that we find similar results when we use probit and logit regressions.

where i and t represent borrower and time, respectively.

In this regression, T_t represents a vector of dummy variables for month t. It begins from 2019-04 and ends in 2021-02. The benchmark time is the beginning of the quarantine in March 2020. By incorporating the interaction term between $RelationshipDummy_i$ and T_t, we aim to elucidate the nuanced dynamics of relationship banking amidst the COVID-19 lockdown, shedding light on its implications for borrower default behavior.

First, we examine whether a parallel trend exists in the default behavior of borrowers with prior relationships and those without, before the inception of the lockdown in March 2020. Figure 10 indicates that all coefficients on the interaction term for the period after June 2019 and prior to March 2020 are statistically insignificant, confirming the parallel trend assumption.

[Insert Figure 10 About Here]

The graph also reveals an interesting trend amid the overall increase in loan defaults during the three months following the lockdown: borrowers with prior relationships with the bank exhibit significantly lower default rates compared to those without such relationships. This finding suggests that maintaining pre-existing relationships may mitigate default risk during periods of economic uncertainty and financial distress.

7.2. The Role of Wealth in Maintaining Banking Relationship

We demonstrate that borrowers tend to pay mortgage installments to preserve the relationship that they have built with the bank. The value of the relationship may depend on its reusability. Hence, borrowers who are more likely to use their banking relationship in the future may have a greater incentive to avoid default. In this context, we investigate the role of borrowers' wealth in loan performance, comparing those with a prior relationship to those without one. On the one hand, wealthier borrowers may place less value on the banking relationship because they possess strong "hard" information that can support their future credit applications (Liberti and Petersen, 2019). On the other hands, such borrowers may attach more value to the relationship because they are more likely to use it in the future, given they may have a greater and wider range of opportunities.

Our aim is to examine the value of relationship among borrowers with different levels of wealth.[26] We estimate the following regression model for our analysis using logit model:[27]

$$\begin{aligned} P(Default)_i = {} & \alpha_i + \beta_1 \, RelationshipDummy_i + \beta_2 \, HomePrice_i \\ & + \beta_3 (RelationshipDummy_i \times HomePrice_i) + \beta_4 \, Characteristics_i + \beta_5 \, RiskMeasures_i + \epsilon_i \end{aligned} \tag{7}$$

[26] The analysis is limited to mortgage loan data for Tehran because home price data is only available for this city.

[27] We also estimate the model using LPM and probit regressions; however, for simplicity, we report only the results of logit regression. We confirm that the results are similar when using LPM and probit regressions.

where *i* represents the borrower.

We define *Default* as a function of a dummy variable indicating prior relationships with the bank ($RelationshipDummy_i$), home price (*HomePrice*), the interaction term between these two variables ($RelationshipDummy_i \times HomePrice_i$), and the control variables used in our main analysis (represented by $Characteristics_i$ and $RiskMeasures_i$). The interaction term is our variable of interest.

Figure 11 display the results. The marginal effects of prior relationships at different home prices are statistically significant at the 1% level (please refer to Appendix A5 for more details).[28] The results indicates that borrowers with higher wealth – those in more expensive neighborhoods – have stronger incentives to avoid default. The marginal effects range from -2.85% to -9.31%. This implies that the default probability of the borrowers – with a prior relationship with the bank – living in a neighborhood with an average home price per square meter of $250 ($1,450) is 2.85% (9.31%) less than the default probability of the borrowers without such a relationship living in the same neighborhood.

[Insert Figure 11 About Here]

Overall, the analysis reveals a geographic disparity in the borrowers' incentive channel. Borrowers residing in Tehran's wealthier districts (the upper regions of the city as shown in Figure 12) demonstrate a stronger response to the potential benefits of maintaining a relationship with the bank. This reinforces the notion that the incentive to avoid default is heterogeneous across wealth levels.

[Insert Figure 12 About Here]

7.3. The Role of Religiosity in Maintaining Banking Relationship

In this subsection, we explore the role of borrowers' religiosity in influencing their incentives to maintain their prior relationships with the bank. According to the literature, more religious individuals are often found to be more socially oriented (Ellison, 1992; Dunfield, 2014; Anadriani and Sabatini, 2015; Tian, 2022), placing a higher value on personal interactions and community connections. This social inclination can enhance personal interactions and accordingly strengthen the relationship with their bank.

To analyze the effect of religiosity on borrowers' incentives to avoid default, we utilize a proxy measure (mosque density) for religiosity similar to Okulicz-Kozaryn (2015) and Cau et al.

[28] Additionally, the difference between the marginal effects at different home prices are statistically significant at the 1% level.

(2018),[29] as direct data on borrowers' religious practices is unavailable. Specifically, we calculate the number of mosques inside circles with radii of 200m, 300m, 400m, 500m, and 600m around the center of each borrower's neighborhood,[30] identified by their 5-digit ZIP code (please see Appendix A6 for summary statistics). Similar to subsection 6.2.1, we use Google Maps to identify the coordination of mosques in Tehran. We hypothesize that higher levels of borrowers' religiosity are correlated with higher density of mosques. This assumption is based on the idea that individuals living nearer to mosques may have more frequent opportunities for participation in religious activities, thereby reinforcing their religious values.

As we need the data of borrowers' residence neighborhood, our analysis is limited to the city of Tehran, where we access the borrowers' residence 5-digit ZIP code neighborhood. In Iran, particularly in Tehran, Islam is the predominant religion, encompassing 99.08% of the population, according to the *National Population and Housing Census (2016)* from the Statistics Center of Iran.[31] Given this demographic context, our analysis focuses exclusively on Muslim borrowers and examines mosques as the central religious institution.

There are 613 mosques in Tehran which are more concentrated in southeastern parts of the city. In our context, given our proxy for religiosity, the people residing in south-eastern are expected to be more religious. Figure 13 shows the distribution of mosques in Tehran.

[Insert Figure 13 About Here]

We examine the effect of religiosity on borrowers' incentives to avoid default by estimating the following regression:[32]

$$\begin{aligned} \mathrm{P}(Default)_i = \; & \alpha_i + \beta_1\, RelationshipDummy_i + \beta_2\, NM_i \\ & + \beta_3 (RelationshipDummy_i \times NM_i) + \beta_4\, Characteristics_i + \beta_5\, RiskMeasures_i + \epsilon_i \end{aligned} \tag{8}$$

where NM_i denotes the number of mosques inside circles with radii of 200m, 300m, 400m, 500m, and 600m around the center of each borrower's neighborhood and i represents the borrower.

The results are reported in columns (1) to (5) Table 17 for NM_i with radii of 200m to 600m, respectively. They show that the coefficient on NM is not significant, which implies that

[29] Okulicz-Kozaryn (2015) uses church density as an indicator of religiosity, and Cau et al. (2018) employ mosque density as a proxy for the strength of local religious norms, values, and institutions.

[30] We ignore radius of 100*m* because the number of mosques inside the 100*m* circles were almost zero for the majority of borrowers (about 96% of the sample).

[31] https://old.sci.org.ir/english/Population-and-Housing-Censuses

[32] In this subsection, we only report the LPM results the sake of brevity. However, we confirm that we find similar results for probit and logit estimation.

religiosity does not affect default risk by itself. However, the findings underscore the significance of religiosity in influencing borrowers' incentives to maintain their banking relationships. The coefficient on the interaction term ($RelationshipDummy_i \times NM_i$) exhibits a negative value across all columns, with statistical significance for circles with radii of 200m to 500m, reported in columns (1) to (4), respectively. This result aligns with Baele et al. (2014), who find that default rates are lower during Ramadan for pious borrowers, likely due to strengthened social relationships and the heightened value of personal interactions. Interestingly, the economic significance declines with the increase in radii of circle around the location of a borrower. Specifically, the addition of a mosque to the neighborhood of a borrower within 200m to 600m from borrowers' residence, is associated with respectively a 1.43%, 1.21%, 0.769%, 0.437%, and 0.145% decline in the probability of default. This finding suggests that as religiosity increases – represented by an increase in the density of mosques – the effect of prior relationships on avoiding loan default becomes more pronounced. It implies that borrowers' incentives to maintain relationships with the bank increases with religiosity. This result aligns with our expectations, highlighting that more religious borrowers attach more value to banking relationship.

[Insert Table 17 About Here]

7.4. The Role of Gender in Maintaining Banking Relationship

The extant literature demonstrates that women face discrimination in access to credit (Bellucci et al., 2010; Alesina et al., 2013; Ongena and Popov, 2016), despite evidence showing no higher delinquency rates among females (Beck et al., 2018). Abedifar et al. (2024) finds that prior banking relationships benefit female borrowers more than male borrowers in obtaining new credit. This posits an important question: do female borrowers have greater incentives to avoid default compared to male borrowers?

This question is worth exploring because, on the one hand, banking relationships may be more valuable for female borrowers than male borrowers. On the other hand, it can be argued that nurturing banking relationships might be even more important for male borrowers, as they are more likely to engage in business activities as part of their long-term career and financial goals (Hira and Mugenda, 2000). Consequently, male borrowers might be more inclined to maintain their banking relationships. Indeed, major financial milestones such as home ownership, business investments, or large-scale purchases typically require solid credit histories. Men aiming for these goals are more likely to preserve their creditworthiness to secure favorable loan terms and interest rates. Furthermore, maintaining a robust credit profile enhances their future financial leverage, granting access to better financial opportunities. This approach helps them capitalize on future prospects and minimize the risks associated with poor credit, thereby strengthening their commitment to avoid default. Given these contrasting viewpoints, a straightforward answer is not apparent and requires further analysis. Therefore, in this subsection, we examine this research question.

We estimate Equation (6) using LPM, probit, and logit regressions. The model includes the explanatory variables of the Equation (1) and the interaction term between the $RelationshipDummy_i$ and gender ($Male_i$). In addition, one might argue that males have more wealth and income, which enable them to better maintain their relationships and default less. Therefore, we control for neighborhood home prices as a proxy for wealth.[33] The model specification is as follows:[34]

$$P(Default)_i = \alpha_i + \beta_1\, RelationshipDummy_i + \beta_2\, Male_i$$
$$+ \beta_3(RelationshipDummy_i \times Male_i) + \beta_4\, Characteristics_i + \beta_5\, RiskMeasures_i + \epsilon_i \quad (9)$$

where *i* represents the borrower.

The estimation results using LPM, probit, and logit regressions are presented in columns (1), (2) and (3) of Table 18, respectively. The coefficient on $RelationshipDummy_i \times Male_i$ is significantly negative in all three columns at the 5% level, which indicates that the association of prior banking relationships and avoiding mortgage default is more pronounced for male borrowers. Specifically, prior banking relationships are associated with a lower default probability for male borrowers by 5.37%, 5.51%, and 5.56%, respectively.[35] However, for female borrowers, the impact is less substantial, with a decrease in default probability of 3.08%, 3.19%, and 3.15%, respectively. The results on control variables are similar to our previous findings.

[Insert Table 18 About Here]

Overall, the results suggests that male borrowers with prior relationship have stronger incentives – compared to female borrowers – to fulfill their mortgage payment obligations, and thereby preserve their banking relationship.

7.5. The Role of Marital Status in Maintaining Banking Relationship

A borrower's incentive to honor a payment claim might depend on her marital status. The potential consequences of defaulting on a loan extend to borrowers' spouse and dependents. This extension could encourage them to fulfill their financial commitments better than single

[33] Similar to subsection 6.2.2, our analysis is limited to mortgage loan data for Tehran.

[34] We verify that there is sufficient variation and observations in *RelationshipDummy* across the male and female sub-samples. Furthermore, the low correlation of 0.018 between *RelationshipDummy* and the gender variable (*Male)* ensures us that our results are not affected by multi-collinearity problem. The results are available upon request.

[35] The effect of prior relationships on the probability of default for male borrowers is the sum of the marginal effects of *RelationshipDummy* and the interaction term between *RelationshipDummy* and *Male* (*RelationshipDummy* × *Male*), while the effect of prior relationships for female borrowers is equal to the marginal effect of *RelationshipDummy*.

borrowers. In addition, married individuals are more likely to engage in long-term financial planning – such as saving their children's education, or preparing for retirement – which requires keeping their banking relationships in good standing to facilitate access to credit. Our investigation shows that married borrowers with prior banking relationship is less likely to default than single borrowers with prior relationship. Therefore, one can argue that married borrowers might have more incentives to avoid default and nurture their banking relationships.

In this subsection, we examine this argument by estimating Equation (10) using LPM, probit, and logit. The dependent variable and explanatory variables are the same as the one in Equation (1). We merely add the interaction terms between $RelationshipDummy_i$ and $Married_i$ to the model as our variable of interest:

$$\begin{aligned} \mathrm{P}(Default)_i &= \alpha_i + \beta_1\, RelationshipDummy_i + \beta_2\, Married_i \\ &\beta_3(RelationshipDummy_i \times Married_i) + \beta_4\, Characteristics_i + \beta_5\, RiskMeasures_i + \epsilon_i \end{aligned} \tag{10}$$

where i represents the borrower.

Table 19 presents the estimation results of Equation (10)[36] using LPM, probit and logit regressions. The results – reported in columns (1), (2), and (3) – show that there is no significance difference in the impact of prior relationships on loan performance between single and married borrowers.

[Insert Table 19 About Here]

8. Concluding Remarks

This study investigates the role of borrowers' prior banking relationship in loan repayments. Using a comprehensive dataset of around 149,230 mortgage loans, we find robust evidence that borrowers with prior banking relationships exhibit significantly lower propensity to default on their loan installments compared to those without such relationships. The results indicate that the borrowers' incentives channel plays a significant role in relationship banking.

Specifically, our LPM, probit, and logit estimates indicate that borrowers maintaining savings or current accounts prior to loan origination have an approximately 4% lower incentive to default, after controlling for an array of borrowers' characteristics and risk factors. This effect persists across various robustness checks, including alternative estimation methods, different

[36] Similar to Equation (9), we verify that there is sufficient variation and observations in *RelationshipDummy* across the married and single sub-samples. Furthermore, the low correlation of -0.06 between *RelationshipDummy* and the marital status variable (*Married*) ensures us that our results are not affected by multi-collinearity problem. The results are available upon request.

default definition, and measures of relationship duration, underscoring the pivotal role of relationship banking in mitigating default risk.

Furthermore, our analysis of the COVID-19 pandemic lockdown as an exogenous shock reveals that borrowers with established banking relationships are less inclined to default on their payment obligations, potentially driven by a desire to preserve valuable relationships that could facilitate future access to financial support. This finding highlights the stabilizing influence of relationship banking during economic crises.

Our study also demonstrates the heterogeneity in borrowers' incentives to avoid default across different wealth levels. Using home prices as a proxy for wealth, we find that more affluent borrowers with prior banking relationships have greater incentives to honor their mortgage payments and thereby preserve their banking relationships possibly because they are more likely to apply for credit in the future. Furthermore, we find that the religiosity strengthens the influence of prior relationships on avoiding defaults.

In addition, our study further extends the literature by showing that borrowers' behavior under a mortgage loan varies across genders. We find that the association of prior banking relationships and incentives to avoid default is stronger for male borrowers relative to female borrowers. These findings highlight the importance of considering demographic factors in understanding the dynamics of borrowers' behavior.

In essence, this study underscores the significance of relationship banking as a powerful mechanism for aligning borrowers' incentives and enhancing loan repayment discipline. By nurturing long-term relationships, banks can cultivate a sense of commitment and trust among borrowers, ultimately reducing default risk and promoting financial stability. These insights carry significant implications for both lending institutions and policymakers, underscoring the importance of fostering enduring bank-borrower relationships as one of the fundamental aspects of a resilient financial system. Banks may consider assigning considerable importance to borrowers' incentives alongside screening and monitoring when assessing loan applications. Moreover, they may choose to prioritize extending loans to customers with longer-established relationships, recognizing the decreased default risk associated with such clients.

References

Abedifar, Pejman, Seyed Javad Kashizadeh, and Steven Ongena (2024). Flood, farms and credit: The role of branch banking in the era of climate change. *Journal of Corporate Finance,* 85, 102544.

Agarwal, Sumit, and Robert Hauswald (2010). Distance and private information in lending. *The Review of Financial Studies*, 23(7), 2757-2788.

Agarwal, Sumit, Souphala Chomsisengphet, Chunlin Liu, Changcheng Song, and Nicholas S Souleles (2018). Benefits of relationship banking: Evidence from consumer credit markets. *Journal of Monetary Economics*, 96, 16–32.

Alesina, Alberto F, Francesca Lotti, and Paolo Emilio Mistrulli (2013). Do women pay more for credit? Evidence from Italy. *Journal of the European Economic Association*, 11 (suppl_1), 45–66.

Andriani, Luca, and Fabio Sabatini (2015). Trust and prosocial behaviour in a process of state capacity building: the case of the Palestinian territories. *Journal of Institutional Economics*, 11(4), 823-846.

Baele, Lieven, Moazzam Farooq, and Steven Ongena (2014). Of religion and redemption: Evidence from default on Islamic loans. *Journal of Banking & Finance*, 44, 141–159.

Bartke, Stephan, and Reimund Schwarze (2008). Risk-averse by nation or by religion? Some insights on the determinants of individual risk attitudes. *Some Insights on the Determinants of Individual Risk Attitudes (October 2008). SOEPpaper, (131).*

Beck, Thorsten, Patrick Behr, and Andreas Madestam (2018). Sex and credit: Do gender interactions matter for credit market outcomes? *Journal of Banking & Finance*, 87, 380–396.

Bedayo, Mikel, Gabriel Jiménez, José-Luis Peydró, and Raquel Vegas Sánchez (2020). Screening and loan origination time: lending standards, loan defaults and bank failures. *Banco de Espana Working Paper*.

Bellucci, Andrea, Alexander Borisov, and Alberto Zazzaro (2010). Does gender matter in bank–firm relationships? Evidence from small business lending. *Journal of Banking & Finance*, 34 (12), 2968–2984.

Berlin, Mitchell, Loretta J. Mester (1999), Deposits and Relationship Lending. *The Review of Financial Studies*, Volume 12, Issue 3, Pages 579–607.

Bonfim, Diana, Ralph De Haas, Alexandra Matyunina, and Steven Ongena (2023). Bank Specialization in Lending to New Firms, *SSRN Electronic Journal, 2014, matyunina.com/wp-content/uploads/2023/10/JMP_Matyunina_Bank_Specialization_StartUps.pdf*.

Boot, Arnoud (2000), Relationship Banking: What Do We Know? *Journal of Financial Intermediation*, vol. 9, issue 1, p. 7-25.

Bradley, Andrew P (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern Recognition* 30, no. 7: 1145-1159.

Breiman, Leo (2001). Random forests. *Machine learning* 45: 5-32.

Bruce, Carol, Maeve E Gearing, Jill DeMatteis, Kerry Levin, Timothy Mulcahy, Jocelyn Newsome, and Jonathan Wivagg (2022). Financial vulnerability and the impact of COVID-19 on American households. *Plos One*, 17 (1), e0262301.

Bruhn, Miriam, and Inessa Love (2014). The real impact of improved access to finance: Evidence from Mexico. *The Journal of Finance*, 69(3), 1347-1376.

Burgess, Robin, and Rohini Pande (2005). Do rural banks matter? Evidence from the Indian social banking experiment. *American Economic Review*, 95(3), 780-795.

Cao, Xun, Haiyan Duan, Chuyu Liu, and Yingjie Wei (2018). Local religious institutions and the impact of interethnic inequality on conflict. *International Studies Quarterly*, 62(4), 765-781.

Chambers, R, P Kokic, P Smith, and M Cruddas (2000). Winsorization for identifying and treating outliers in business surveys. In *Proceedings of the second international conference on establishment surveys* (Vol. 1, pp. 1-12). *American Statistical Association.*

Chen, Tianqi, and Carlos Guestrin (2016). Xgboost: A scalable tree boosting system. *In Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining*, pp. 785-794.

Ciampi, Francesco, and Niccolò Gordini (2013). Small enterprise default prediction modeling through artificial neural networks: an empirical analysis of Italian small enterprises. *Journal of Small Business Management*, 51(1): 23-45.

Cleves, Mario, William Gould, Roberto Gutierrez, and Yulia V Marchenko (2008). An introduction to survival analysis using Stata. *Stata Press*.

Cox, David R (1972). Regression models and life-tables. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34 (2), 187–202.

Davis, Jesse, and Mark Goadrich (2006). The relationship between Precision-Recall and ROC curves. *In Proceedings of the 23rd international conference on Machine learning*, pp. 233-240.

Degryse, Hans, Vasso Ioannidou, and Erik von Schedvin (2016). On the nonexclusivity of loan contracts: An empirical investigation. *Management Science* 62, no. 12: 3510-3533.

Diamond, Douglas W (1984). Financial intermediation and delegated monitoring. *The Review of Economic Studies*, 51 (3), 393–414.

Dohmen, Thomas, Armin Falk, David Huffman, and Uwe Sunde (2012). The intergenerational transmission of risk and trust attitudes. *The Review of Economic Studies*, 79(2), 645-677.

Dunfield, Kristen A (2014). A construct divided: Prosocial behavior as helping, sharing, and comforting subtypes. *Frontiers in Psychology*, 5, 958.

Ellison, Christopher G (1992). Are religious people nice people? Evidence from the National Survey of Black Americans. *Social Forces*, 71(2), 411-430.

Fiordelisi, Franco, Stefano Monferrà, and Gabriele Sampagnaro (2014). Relationship lending and credit quality. *Journal of Financial Services Research*, 46, 295–315.

Fuster, Andreas, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther (2022). Predictably unequal? The effects of machine learning on credit markets. *The Journal of Finance* 77, no. 1: 5-47.

Garmaise, Mark J., and Tobias J. Moskowitz (2006). Bank mergers and crime: The real and social effects of credit market competition. *The Journal of Finance*, 61(2), 495-538.

Grambsch, Patricia M and Terry M Therneau (1994). Proportional hazards tests and diagnostics based on weighted residuals. *Biometrika*, 81 (3), 515–526.

Hilary, Gilles and Kai Wai Hui (2009). Does religion matter in corporate decision making in America?. *Journal of Financial Economics* 93(3), 455–473.

Hira, Tahira K., and Olive Mugenda (2000). Gender differences in financial perceptions, behaviors and satisfaction. *Journal of Financial Planning-Denver*-13, no. 2: 86-93.

Hosmer, David W, Stanley Lemeshow, and Susanne May (2008). Applied survival analysis: regression modeling of time-to-event data. chapter 6, Vol. 618, *John Wiley & Sons*.

Imbens, Guido W and Donald B Rubin (2015). Causal inference in statistics, social, and biomedical sciences. *Cambridge University Press*.

Keys, Benjamin J, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig (2010). Did securitization lead to lax screening? Evidence from subprime loans. *The Quarterly Journal of Economics*, 125 (1), 307–362.

Khandani, Amir E., Adlar J. Kim, and Andrew W. Lo (2010). Consumer credit-risk models via machine-learning algorithms. *Journal of Banking & Finance* 34, no. 11: 2767-2787.

Kuitunen, Ilari, Ville T Ponkilainen, Mikko M Uimonen, Antti Eskelinen, and Aleksi Reito (2021). Testing the proportional hazards assumption in cox regression and dealing with possible nonproportionality in total joint arthroplasty research: methodological perspectives and review. *BMC Musculoskeletal Disorders*, 22 (1), 489.

Levine, Ross, Chen Lin, Qilin Peng, and Wensi Xie (2020). Communication within banking organizations and small business lending. *The Review of Financial Studies*, 33(12), 5750-5783.

Liberti, José María, and Mitchell A. Petersen (2019). Information: Hard and soft. *Review of Corporate Finance Studies* 8, no. 1: 1-41.

Liu, Eric Y (2010). Are risk-taking persons less religious? Risk preference, religious affiliation, and religious participation in Taiwan. *Journal for the Scientific Study of Religion* 49, no. 1: 172-178.

Manove, Michael, A Jorge Padilla, and Marco Pagano (2001). Collateral versus project screening: A model of lazy banks. *Rand Journal of Economics*, pp. 726–744.

Nguyen, Hoai-Luu Q. (2019). Are credit markets still local? Evidence from bank branch closings. *American Economic Journal: Applied Economics*, 11(1), 1-32.

Nigmonov, Asror and Syed Shams (2021). COVID-19 pandemic risk and probability of loan default: evidence from marketplace lending market. *Financial Innovation*, 7 (1), 83.

Noussair, Charles N., Stefan T. Trautmann, Gijs Van de Kuilen, and Nathanael Vellekoop (2013). Risk aversion and religion. *Journal of Risk and Uncertainty*, 47, 165-183.

Okulicz-Kozaryn, Adam (2015). The more religiosity, the less creativity across US counties. *Business Creativity and the Creative Economy*, 1(1), 81-87.

Ongena, Steven and Alexander Popov (2016). Gender bias and credit access. *Journal of Money, Credit and Banking*, 48 (8), 1691–1724.

Ongena, Steven and David C Smith (2001). The duration of bank relationships. *Journal of Financial Economics*, 61 (3), 449–475.

Padilla, A. Jorge, and Marco Pagano (1999). Endogenous communication among lenders and entrepreneurial incentives. *The Review of Financial Studies*, 10(1), 205-236.

Petersen, Mitchell A., and Raghuram G. Rajan (1994). The Benefits of Lending Relationships: Evidence from Small Business Data. *Journal of Finance*, vol. 49:1, pages 3-37.

Puri, Manju, Jörg Rocholl, and Sascha Steffen (2017). What do a million observations have to say about loan defaults? Opening the black box of relationships. *Journal of Financial Intermediation*, 31, 1–15.

Schäfer, Larissa (2019). “Forgive but not forget”: The behavior of relationship banks when firms are in distress. *Review of Finance*, 23 (6), 1079-1114.

Schoar, Antoinette (2012). The personal side of relationship banking. *Available at SSRN 2024653*.

Schoenfeld, David (1982). Partial residuals for the proportional hazards regression model. *Biometrika*, 69 (1), 239–241.

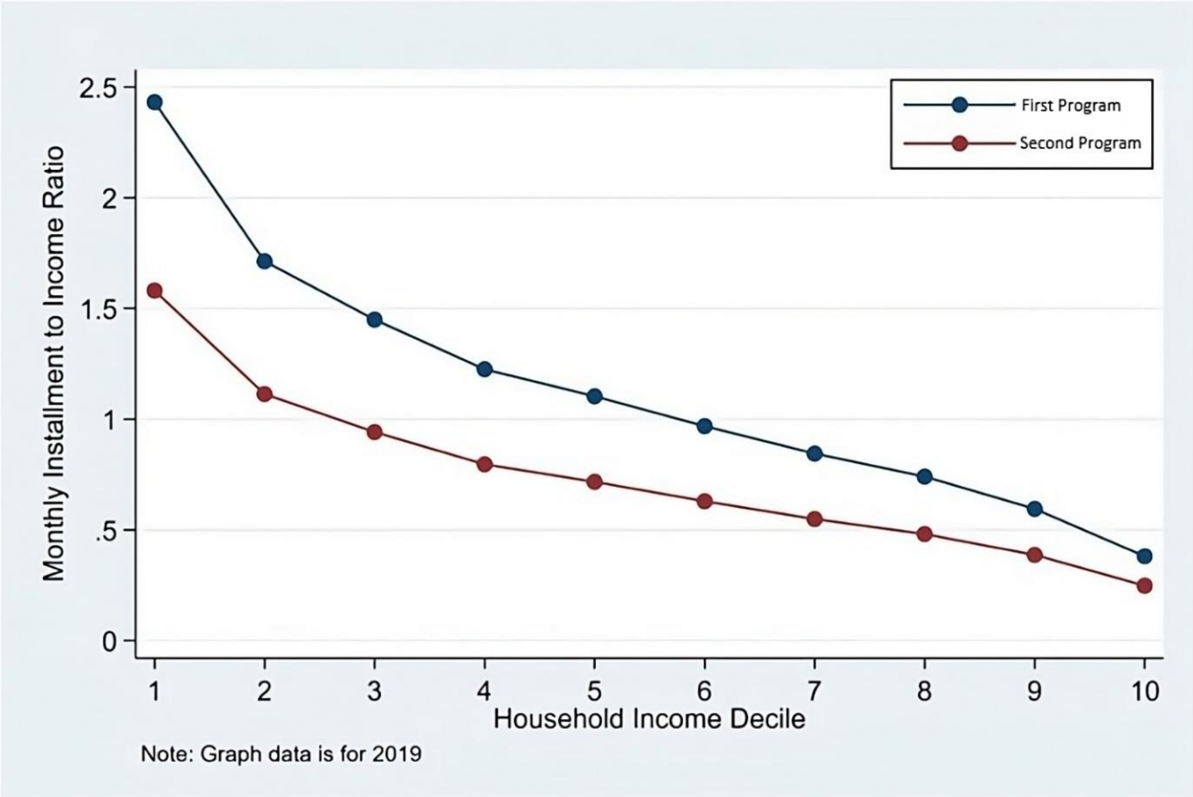
Tian, Ying (2022). Reflection on Religion’s Role in Society: Assessing Religious Function. *International Journal of Frontiers in Sociology*, 4(13).

Figures

Figure 1: Mortgage Papers' Average Price

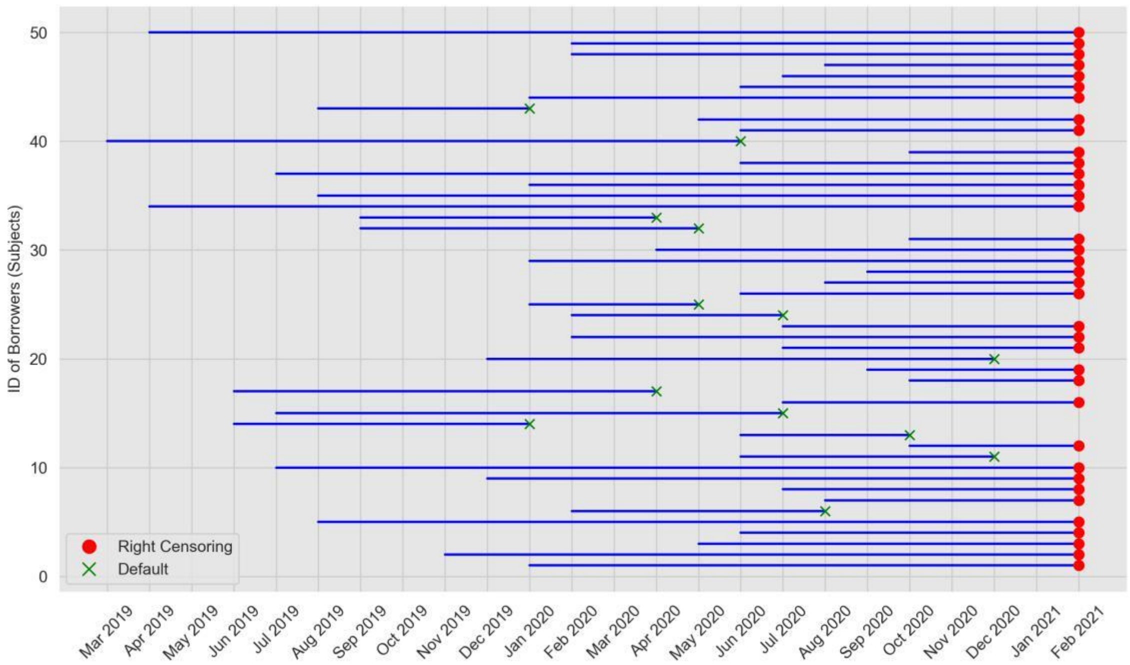
This figure plots the average market price of mortgage papers in $ over the study period from April 2019 to October 2020.

Figure 2: The Installment-to-Income Ratio



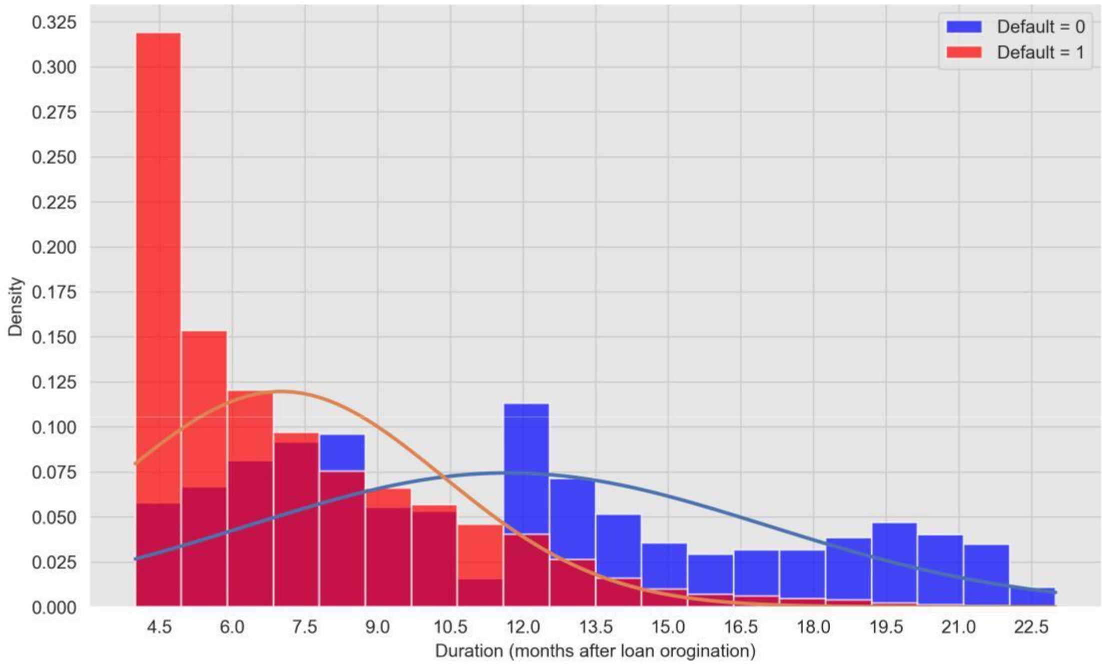
This graph exhibits the average loan installment-to-income ratio (y-axis) across income deciles (x-axis) for households in our sample under the *First Program* and the *Second Program* in 2019.

Figure 3: Time-to-Event and Time-to-Censoring



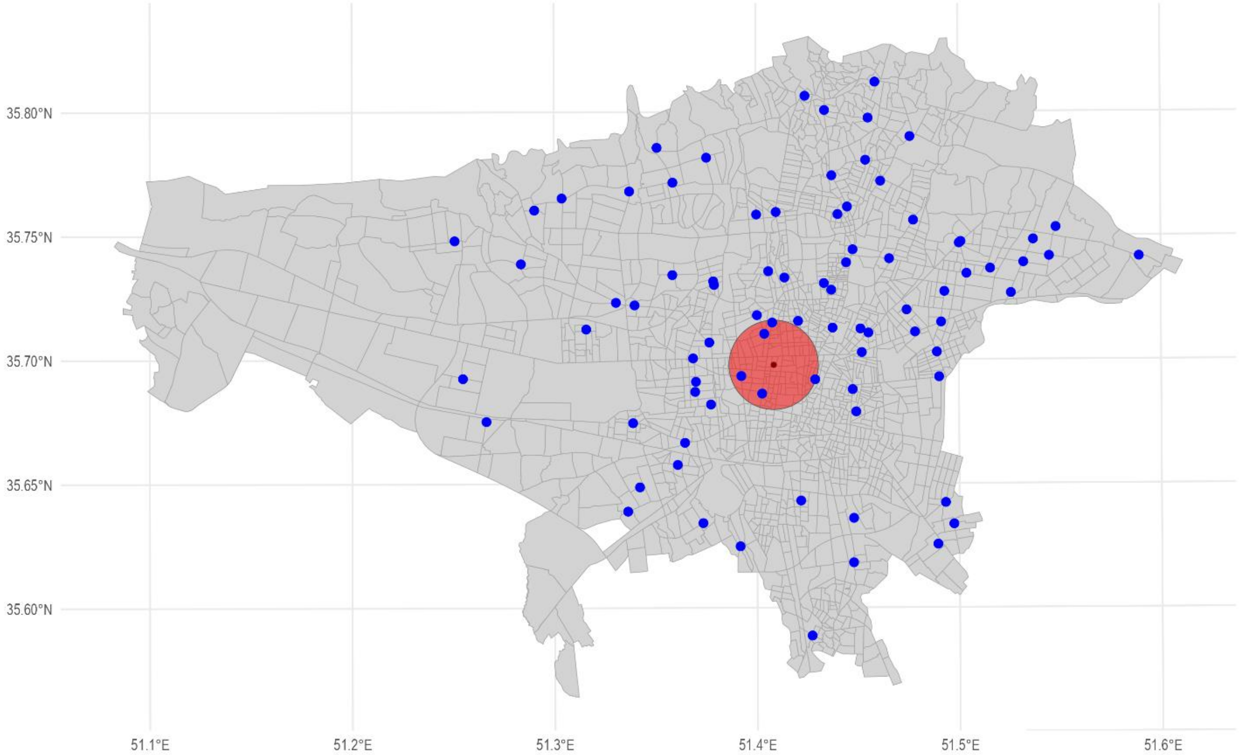
This figure shows the entry time, time-to-event, and time-to-censoring for a 50-borrower random sample from our dataset. Each horizontal line implies one borrower. The red dots at the end of lines indicate that the borrower is censored and the green crosses show that the borrower has left the data by default on her loan.

Figure 4: Histograms of Borrowers' Follow-Up Period



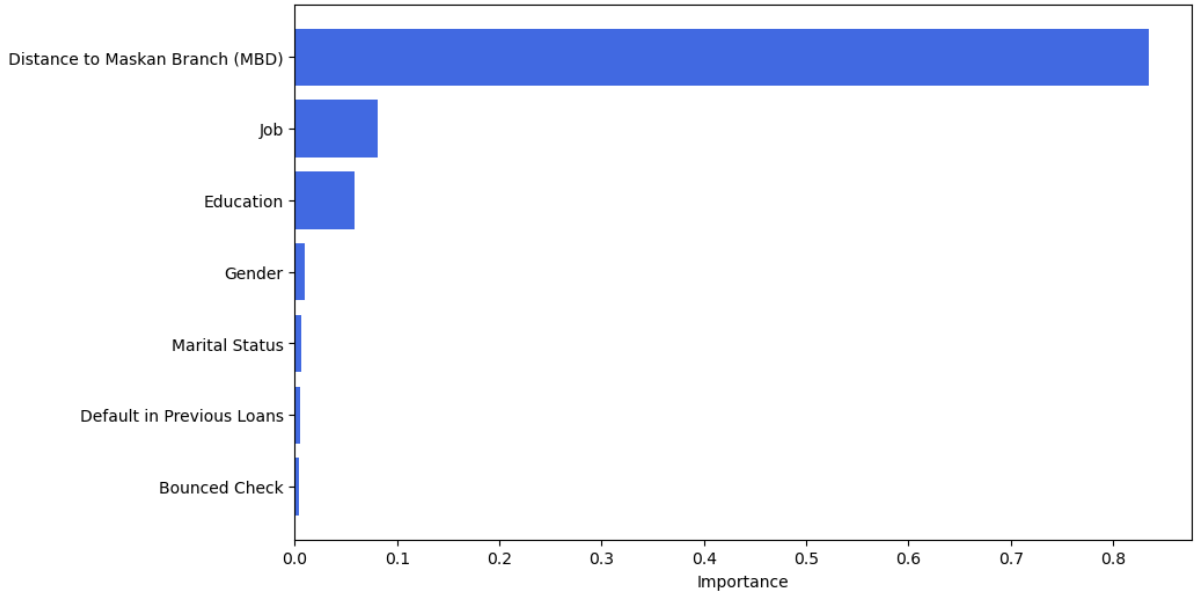
The figure depicts the histograms of the follow-up period for borrowers without default and those with default. The follow-up period is the duration (in months) that a borrower is observed in the data up until they exit the sample either as they default or the right-censoring. The x-axis represents the duration in months, while the y-axis represents the density of borrowers with a specific follow-up period.

Figure 5: Distribution of Maskan Bank Branches in Tehran, with a 2000m Radius Circle Around the Center of a Borrower's 5-Digit ZIP Code Neighborhood.



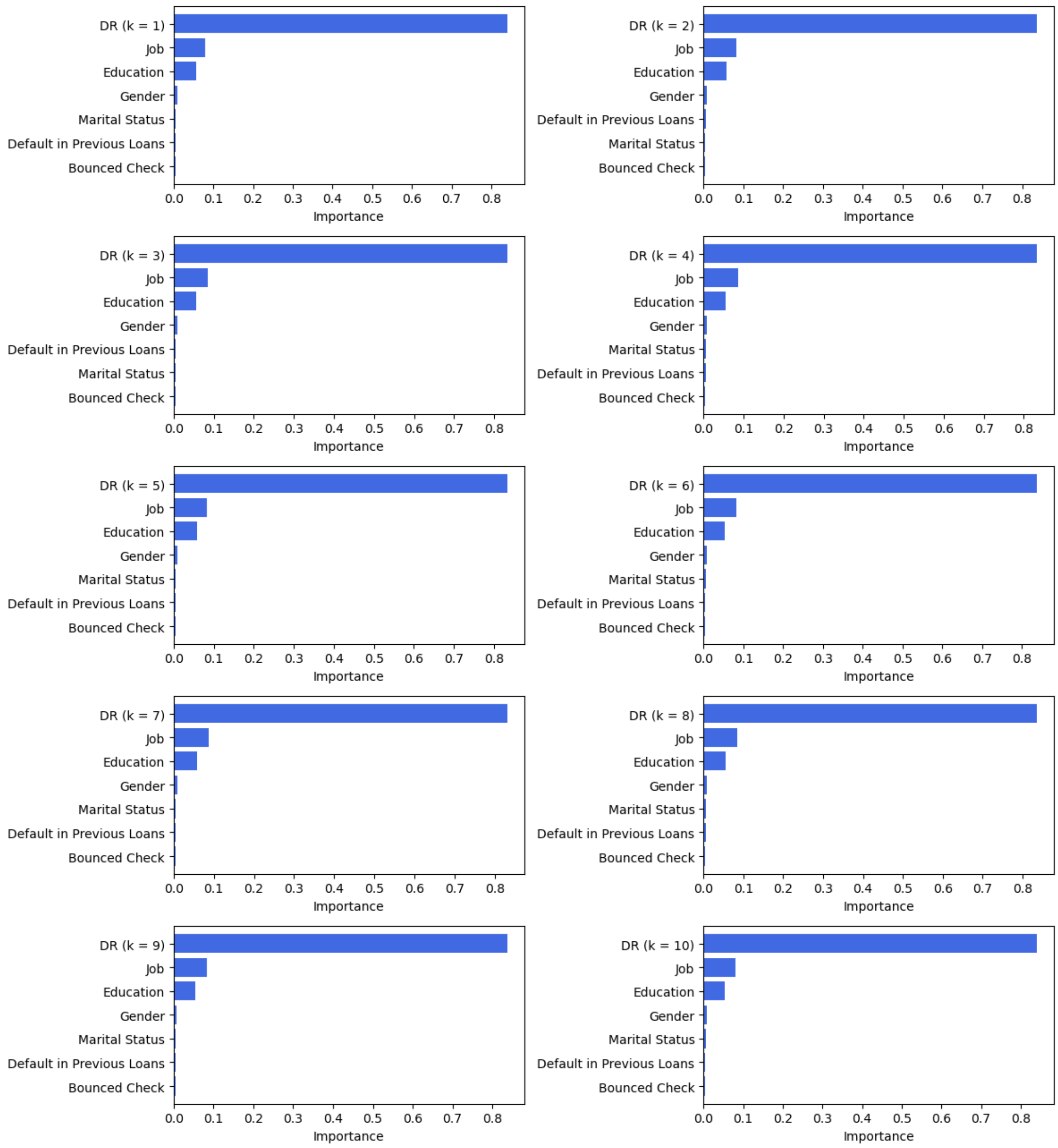
This figure illustrates the distribution of Maskan Bank branches across the map of Tehran. The red circle represents a 2000-meter radius, centered on a borrower's 5-digit ZIP code area. The x-axis depicts the longitude, while the y-axis displays the latitude.

Figure 6: The Results of Random Forest Analysis: The Importance of the Distance to Maskan Branch (MBD) and Borrowers' Characteristics



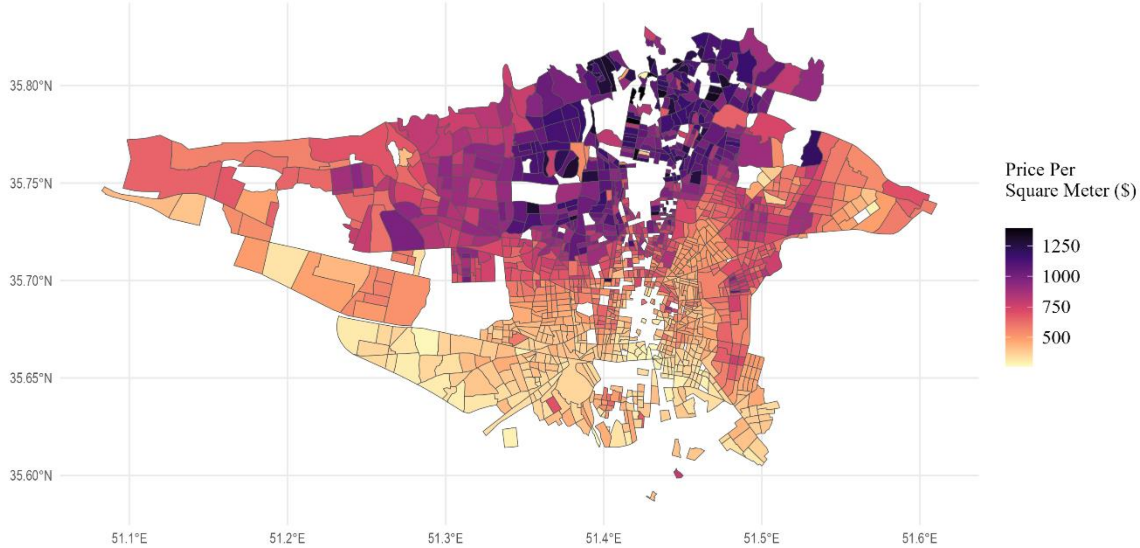
The figure presents the results of the random forest analysis, illustrating the feature importance of various variables in predicting whether a borrower has a relationship with Maskan Bank. The variable of interest is the *Distance to Maskan Branch* (*MBD*). The x-axis represents the percentage of importance for each variable, while the y-axis lists the variables.

Figure 7: The Results of Random Forest Analysis: The Importance of the Distance Ratio (DR) and Borrowers' Characteristics



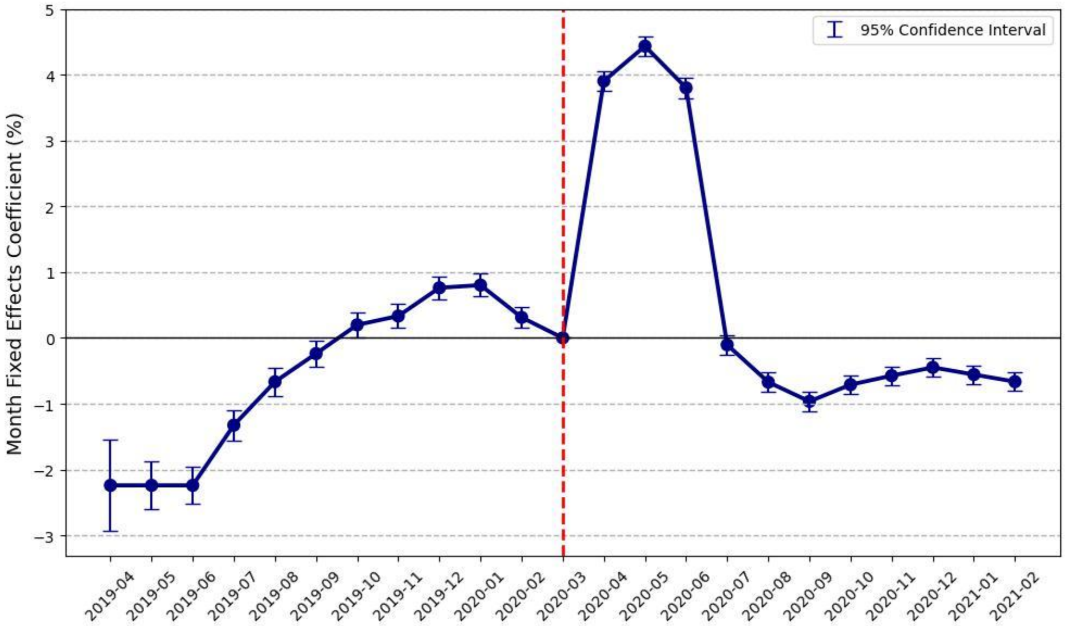
The figure presents the results of the random forest analysis, illustrating the feature importance of Distance Ratio (*DR*) for different values of *k* and borrowers' characteristics in predicting whether a borrower has a relationship with Maskan Bank. The value *k* represents the number of nearest branches of other banks that we use to calculate the denominator of Equation (5). The variable of interest is *DR*. The x-axis represents the percentage of importance for each variable, while the y-axis lists the variables.

Figure 8: Distribution of Home Price in Tehran



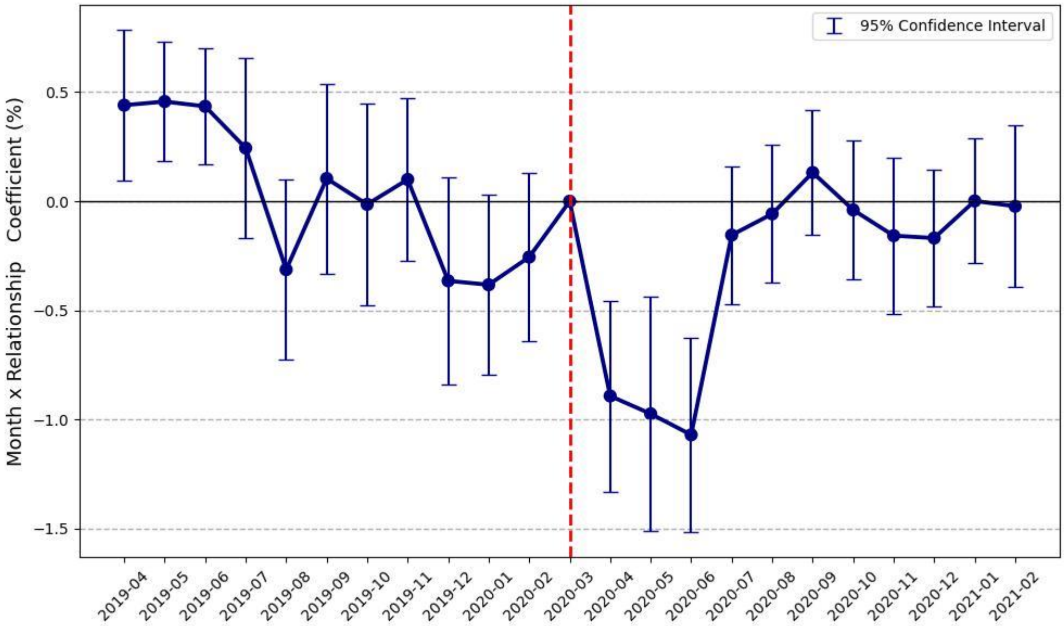
This figure depicts the distribution of home price per square meter in $ across the first 5-digit ZIP code neighborhoods in Tehran, Iran. The x-axis represents longitude and the y-axis indicates latitude. The color of each zone reflects a specific home price, indicated in the legend bar. The districts without color represent missing values.

Figure 9: COVID-19 Lockdown and the Loan Defaults Trend



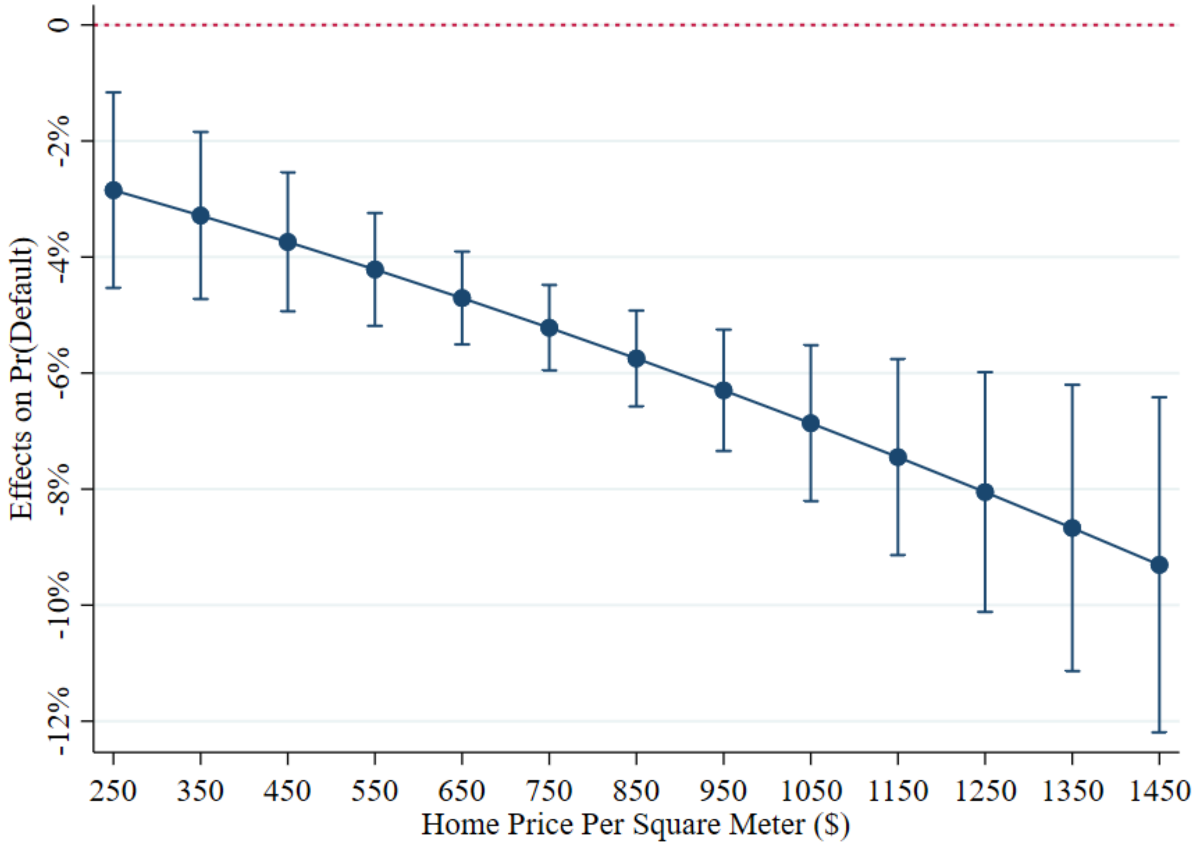
This figure shows the trend of loan defaults before and after the COVID-19 lockdown. The x-axis represents months, and the y-axis indicates the percentage of loan defaults. The vertical lines denote the 95% confidence interval. The benchmark month is March 2020.

Figure 10: The Effect of Prior Relationships in Loan Default During the COVID-19 Lockdown



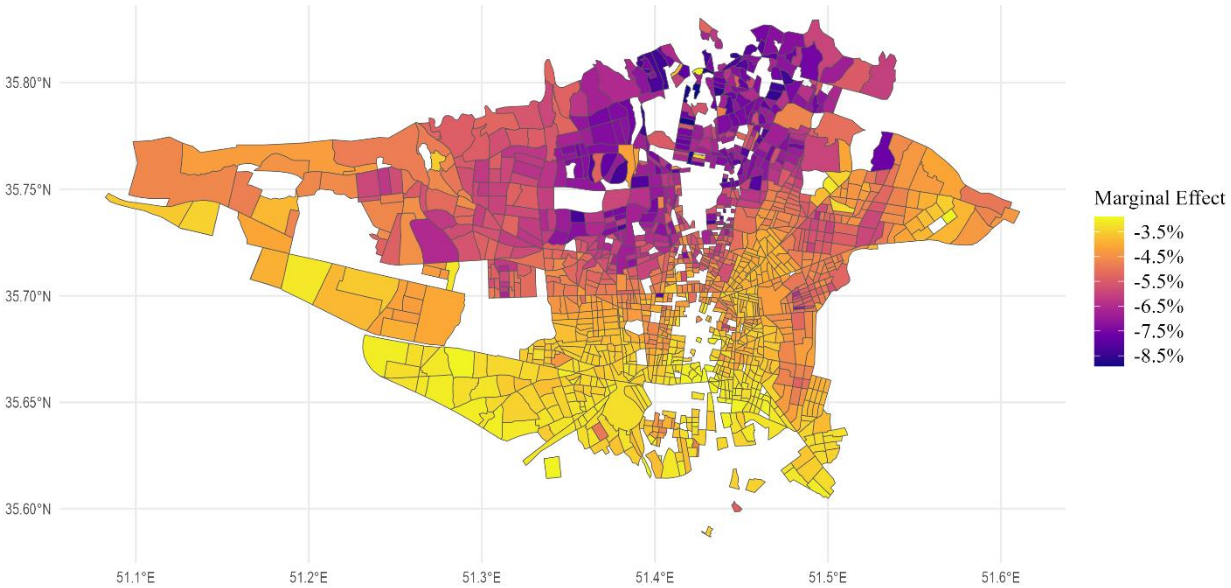
This figure depicts the relationship between a borrower's prior relationships with Maskan Bank and the risk of loan default during the COVID-19 lockdown period. The x-axis represents the month and the y-axis indicates the coefficients on *Month×RelationshipDummy* interaction term resulted from Equation (6). The vertical lines denote the %95 confidence interval. The benchmark month is March 2020.

Figure 11: Marginal Effect of Having Prior Relationships on Probability of Default across Different Home Price Brackets



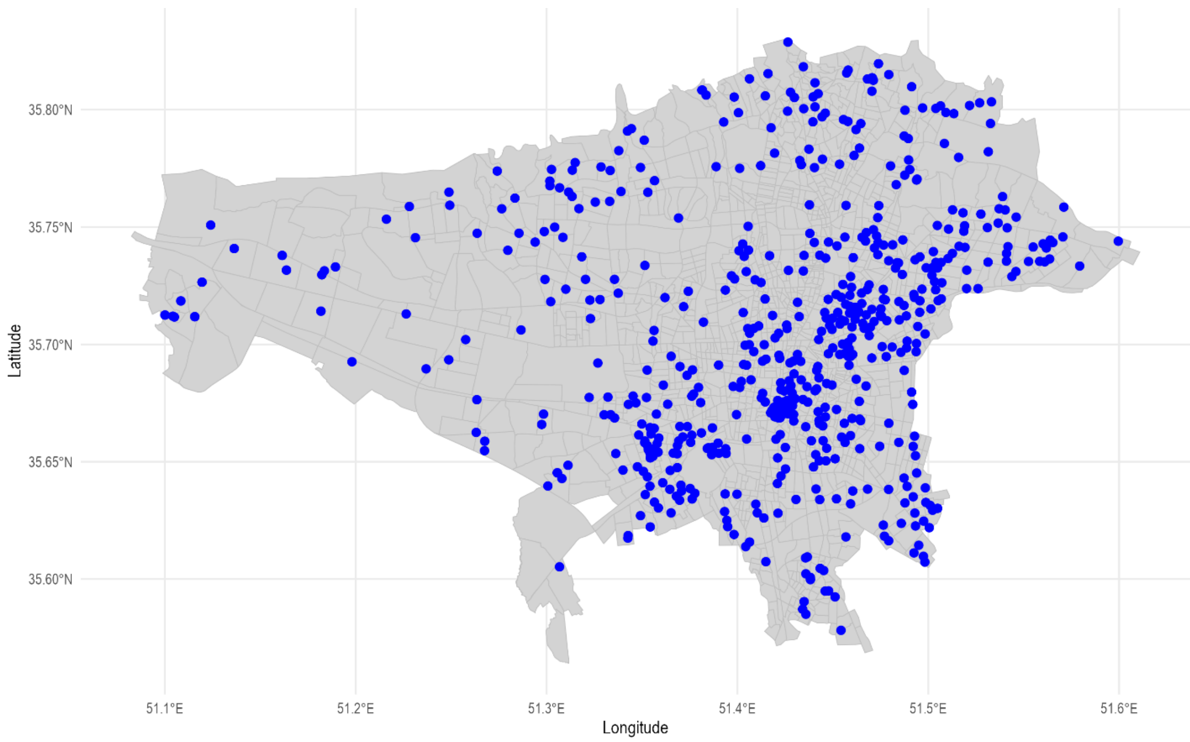
This figure depicts the average marginal effect of having prior relationships with the bank on the probability of loan default across different home prices. The x-axis represents the home price per square meter in $ and the y-axis indicates the change in the probability of default in percentage resulted from Equation (7). The vertical lines denote the 95% confidence interval.

Figure 12: Heterogeneous Effect of Having a Prior Relationship on Probability of Default across Tehran's Zones



This figure depicts the average marginal effect (%) of having prior relationships on the probability of default at each district zone of Tehran. The x-axis represents longitude and the y-axis indicates latitude. The color of each zone reflects a specified marginal effect in percentage, indicated in the legend bar. The districts without color represent missing values.

Figure 13: Distribution of Mosques in Tehran



This figure illustrates the distribution of Mosques across the map of Tehran. The x-axis depicts the longitude, while the y-axis displays the latitude.

Tables

Table 1: Summary Statistics

	Mean	SD	Min	Max	N
3-Month Default (%)	26.49	44.12	0.00	100	149,230
6-Month Default (%)	5.82	23.41	0.00	100	131,784
RelationshipDummy (%)	34.25	47.45	0.00	100	149,230
RelationshipDuration (in Years)	9.33	5.78	0.00	30.66	51,121
Male (%)	58.37	49.29	0.00	100	149,230
Married (%)	86.71	33.94	0.00	100	149,230
Effective Rate (%)	20.42	1.08	17.92	22.97	149,230
Loan to Value (%)	34.56	21.14	4.13	83.33	146,291
Bounced Check Record (%)	3.65	18.76	0.00	100	149,230
Default Record in Previous Loans (%)	12.08	32.59	0.00	100	149,230

This table presents the summary statistics of borrowers' characteristics, risk measures, and loan defaults. The sample includes loans originated by Maskan Bank under the *First Program* of loan issuance from April 2019 to February 2021. Please refer to Appendix A1 for variable definitions.

Table 2: Covariate Balance Test

	Without Prior Relationship (N_{nr} = 96,438)		With Prior Relationship (N_r = 49,853)		Overlap Measures			
							$\pi^{0.05}$	
	Mean	(S.D.)	Mean	(S.D.)	Nor Dif	Log Ratio of STD	Without Prior Relationship	With Prior Relationship
Male	0.58	-0.49	0.6	-0.49	0.04	-0.01	0	0
Married	0.88	-0.32	0.84	-0.37	-0.14	0.14	0	0
Loan to Value	0.35	-0.21	0.32	-0.21	-0.16	-0.01	0.02	0.04
Bounced Check	0.03	-0.17	0.05	-0.22	0.11	0.27	0	0
Default in Previous Loans	0.12	-0.33	0.12	-0.32	-0.01	-0.01	0	0
Effective Rate	0.2	-0.01	0.2	-0.01	0.01	0	0.05	0.05
Multivariate Measures								
P-Score	0.34	(0.05)	0.35	(0.06)	0.24	0.10	0.05	0.06
Linearized P-Score	-0.69	(0.23)	-0.63	(0.25)	0.24	0.08	0.05	0.06

This table presents the overlap assessment of covariates between borrowers with and without prior relationships with the bank. The overlap measures are normalized differences (*Nor Dif*), dispersion difference (*Log Ratio of STD*), and coverage frequency ($\pi^{0.05}$). Additionally, multivariate measures such as propensity scores (*P-Score*) and linearized propensity scores (*Linearized P-Score*) are reported to assess the overall balance between the two groups.

Table 3: Average Marginal Effects of LPM, Probit, and Logit Regressions

	(1) LPM	(2) Probit	(3) Logit
Borrower's Relationship			
RelationshipDummy	-0.0419***	-0.0423***	-0.0429***
	(0.00332)	(0.00324)	(0.00327)
Borrower's Characteristics			
Male	-0.000224	0.000467	0.000286
	(0.00299)	(0.00298)	(0.00301)
Married	0.00246	0.00277	0.00273
	(0.00564)	(0.00578)	(0.00582)
Borrower's Risk Measures			
Loan to Value	0.000513***	0.000514***	0.000500***
	(0.0000763)	(0.0000752)	(0.0000758)
Bounced Check	0.159***	0.140***	0.141***
	(0.0115)	(0.00954)	(0.00948)
Default in Previous Loans	0.509***	0.394***	0.377***
	(0.00679)	(0.00443)	(0.00427)
Effective Rate	-0.0728***	-0.0773***	-0.0760***
	(0.00264)	(0.00283)	(0.00290)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	146,242	146,242	146,242
R^2 (Pseudo R^2)	0.198	0.163	0.162
F (Wald Chi2)	335.37	7887.08	6981.64

This table exhibits the average marginal effects resulted from Equation (1) using the LPM, probit, and logit estimation. The dependent variable is a binary variable equal to 1 if the borrower defaults within the full observation period after loan origination where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is *RelationshipDummy*. We control for *Male, Married, Loan to Value, Bounced Check, Default in Previous Loans,* and *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error terms are clustered at the city level in all models. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 4: Descriptive Statistics of Censored and Failed Subjects

Subject Type	Mean	SD	Min	Max	N
Censored Subjects (Default = 0)	11.60	5.40	3	23	109,706
Failed Subjects (Default = 1)	6.75	3.37	3	23	39,524

This table presents the statistics on censored and failed borrowers within the survival data. A censored borrower refers to one who exits the dataset for reasons other than default. Conversely, a failed borrower is one who exits the dataset due to default.

Table 5: Cox Proportional Hazard Regression

	(1) Hazard Ratio	(2) Hazard Ratio	(3) Hazard Ratio
Borrower's Relationship			
RelationshipDummy	0.818***	0.844***	0.804***
	(0.0126)	(0.0128)	(0.0110)
Borrower's Characteristics			
Male		1.107***	0.983
		(0.0140)	(0.0157)
Married		1.074*	0.993
		(0.0360)	(0.0282)
Borrower's Risk Measures			
Loan to Value			1.002***
			(0.0004)
Bounced Check			1.756***
			(0.0734)
Default in Previous Loans			4.990***
			(0.125)
Effective Rate			0.891***
			(0.0066)
Job	No	Yes	Yes
Education	No	Yes	Yes
N	149,181	149,181	146,242
Wald Chi2	170.0	1367.5	10400.6

This table exhibits the hazard ratios resulted from the Cox proportional hazard model (Equation (2)). The dependent variable is the time to 3-month default indicator. Column (1) includes the *RelationshipDummy* as the explanatory variable. In column (2), we add borrowers' characteristics, i.e. *Male and Married.* Column (3) reports the result when we incorporate borrowers' risk measures, i.e., *Loan to Value, Bounced Check, Default in Previous Loans,* and *Effective Rate*. *Job* and *Education* fixed effects are included in columns (2) and (3). They are jointly significant; however, for brevity, we do not report their coefficients. The error term is clustered at the city level in all models. Standard errors of exponentiated coefficients are in parenthesis. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 6: Default Probability Prediction by Machine Learning Models

	(1) Random Forest	(2) XGBoost	(3) Neural Networks
Borrower's Relationship			
RelationshipDummy	-0.0323***	-0.0395***	-0.0380***
	(0.0024)	(0.0028)	(0.0021)
Borrower's Characteristics			
Male	0.0011	0.0088***	0.0110***
	(0.0030)	(0.0015)	(0.0032)
Married	0.0091**	0.0161***	0.0135***
	(0.0035)	(0.0017)	(0.0035)
Borrower's Risk Measures			
Loan to Value	-0.0003	-0.0086**	0.0287***
	(0.0021)	(0.0035)	(0.0026)
Bounced Check	0.1639***	0.2194***	0.2417***
	(0.0014)	(0.0019)	(0.0033)
Default in Previous Loans	0.4266***	0.5140***	0.5323***
	(0.0023)	(0.0018)	(0.0027)
Effective Rate	-0.0869***	-0.1374***	-0.2634***
	(0.0018)	(0.0035)	(0.0026)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
ROC AUC	0.7316	0.7603	0.7507
Average Precision Score	0.5479	0.5765	0.5663
Brier Score	0.1581	0.1518	0.1526
R^2	0.1795	0.2121	0.2080

This table shows the effect of variables on the probability of default using three machine learning models: Random Forest, XGBoost, and Neural Networks, as presented in columns (1), (2), and (3), respectively. The outcome variable is binary, representing a classification problem where the outcome is one if the borrower defaults within the full observation period after loan origination. Default is defined as failing to repay installments for three consecutive months. The variable of interest is *RelationshipDummy*. We control for *Male, Married, Loan to Value, Bounced Check, Default in Previous Loans*, and *Effective Rate*. We also include *Job* and *Education* as categorical features; however, for brevity, we do not report their effects. The standard errors are computed using bootstrap sampling. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 7: Summary Statistics of the Number of Maskan Bank Branches and Branches of other Banks Inside the Circles with Radii of 500, 1000, 1500, 2000 Meters.

	Mean	SD	Min	Max	N
Maskan Bank Branches					
500m Circle	0.20	0.43	0.00	2.00	28,430
1000m Circle	0.71	0.80	0.00	4.00	28,430
1500m Circle	1.41	1.22	0.00	6.00	28,430
2000m Circle	2.38	1.72	0.00	8.00	28,430
Other Banks Branches					
500m Circle	2.97	4.11	0.00	32.00	28,430
1000m Circle	11.19	11.73	0.00	72.00	28,430
1500m Circle	24.04	22.70	0.00	141.00	28,430
2000m Circle	41.56	37.10	0.00	230.00	28,430

This table presents summary statistics for the number of Maskan Bank branches and other banks located within circles of varying radii – 500, 1000, 1500, and 2000 meters – centered at the borrowers' 5-digit ZIP code neighborhood center.

Table 8: The Role of Geographic Distance to Maskan Bank Branches

	(1) 500*m* Circle	(2) 1000*m* Circle	(3) 1500*m* Circle	(4) 2000*m* Circle
Number of Maskan Branches (*NMB*)	0.0215***	0.0131***	0.0136***	0.0127***
	(0.00656)	(0.00351)	(0.00232)	(0.00163)
Male	0.000158	0.000286	0.000386	0.000855
	(0.00599)	(0.00599)	(0.00598)	(0.00598)
Married	-0.0421***	-0.0418***	-0.0410***	-0.0407***
	(0.00742)	(0.00741)	(0.00742)	(0.00741)
Bounced Check	0.111***	0.111***	0.112***	0.114***
	(0.0190)	(0.0190)	(0.0190)	(0.0190)
Default in Previous Loans	0.0509***	0.0509***	0.0517***	0.0530***
	(0.0150)	(0.0150)	(0.0150)	(0.0150)
Job	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes
N	28,430	28,430	28,430	28,430
Pseudo R^2	0.0121	0.0122	0.0127	0.0134
Wald Chi2	437.16	439.85	459.16	483.96

This table presents the marginal effects of the number of Maskan Bank branches located inside circles with varying radii – 500, 1000, 1500, and 2000 meters – centered at the neighborhood center of borrowers' 5-digit ZIP codes on having relationship with Maskan Bank. The dependent variable is a binary variable that takes the value of one if the borrower has a relationship with Maskan Bank and zero otherwise. The variable of interest is *Number of Maskan Branches* (*NMB*). We control for *Male, Married, Bounced Check,* and *Default in Previous Loans*. We also include *Job* and *Education* fixed effects. The fixed effects are jointly significant; however, for brevity, we do not report their coefficients. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 9: Summary Statistics of the Average Distance between the Borrowers' 5-digit ZIP Code Neighborhood Center and *k* Nearest Maskan Bank and Other Banks Branches

	Mean	SD	Min	Max	N
Maskan Bank Branches					
Dist. to the Nearest	1,287.3	1,460.6	68.7	14,318.9	28,430
Avg. Dist. to the 2 Nearest	1,635.5	1,473.4	219.7	14,471.9	28,430
Avg. Dist. to the 3 Nearest	1,952.7	1,534.0	464.0	15,023.9	28,430
Avg. Dist. to the 4 Nearest	2,232.8	1,592.5	671.8	15,532.7	28,430
Avg. Dist. to the 5 Nearest	2,486.5	1,653.4	942.2	16,047.5	28,430
Avg. Dist. to the 6 Nearest	2,723.0	1,725.6	1,106.6	16,611.2	28,430
Avg. Dist. to the 7 Nearest	2,942.3	1,787.2	1,328.2	17,074.1	28,430
Avg. Dist. to the 8 Nearest	3,147.0	1,852.1	1,489.2	17,586.5	28,430
Avg. Dist. to the 9 Nearest	3,337.7	1,914.3	1,574.6	18,077.2	28,430
Avg. Dist. to the 10 Nearest	3,519.9	1,970.3	1,728.1	18,512.6	28,430
Other Banks Branches					
Dist. to the Nearest	474.5	354.3	13.4	3,519.0	28,430
Avg. Dist. to the 2 Nearest	543.7	378.8	28.0	3,571.4	28,430
Avg. Dist. to the 3 Nearest	609.3	419.0	33.4	5,055.1	28,430
Avg. Dist. to the 4 Nearest	666.5	455.8	56.5	5,954.4	28,430
Avg. Dist. to the 5 Nearest	715.7	485.1	70.6	6,504.1	28,430
Avg. Dist. to the 6 Nearest	760.9	507.5	91.0	6,871.1	28,430
Avg. Dist. to the 7 Nearest	802.6	526.4	105.6	7,136.5	28,430
Avg. Dist. to the 8 Nearest	843.4	542.1	114.2	7,345.2	28,430
Avg. Dist. to the 9 Nearest	881.9	556.2	117.6	7,509.8	28,430
Avg. Dist. to the 10 Nearest	919.4	568.9	124.4	7,652.7	28,430

This table shows the summary statistics of the average distance between the borrowers' 5-digit ZIP code neighborhood center and *k* nearest Maskan Bank and other banks branches. The values are measured in meter.

Table 10: The Marginal Effects of Distance to Maskan Branch (MBD) on Having Relationship with Maskan Bank

	(1) Nearest Branch	(2) 2 Nearest Branches	(3) 3 Nearest Branches	(4) 4 Nearest Branches	(5) 5 Nearest Branches	(1) 6 Nearest Branches	(2) 7 Nearest Branches	(3) 8 Nearest Branches	(4) 9 Nearest Branches	(5) 10 Nearest Branches
Distance to Maskan Branch (MBD)	-0.0019***	-0.0020***	-0.0021***	-0.0020***	-0.0019***	-0.0019***	-0.0019***	-0.0018***	-0.0018***	-0.0018***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Male	0.0006	0.0008	0.0009	0.0010	0.0011	0.00118	0.00125	0.00134	0.00140	0.00145
	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.0059)	(0.00598)	(0.00598)	(0.00598)	(0.00598)	(0.00598)
Married	-0.0404***	-0.0400***	-0.0397***	-0.0394***	-0.0394***	-0.0393***	-0.0392***	-0.0391***	-0.0390***	-0.0390***
	(0.0074)	(0.0074)	(0.0074)	(0.0074)	(0.0074)	(0.00740)	(0.00740)	(0.00740)	(0.00740)	(0.00740)
Bounced Check	0.111***	0.112***	0.112***	0.112***	0.113***	0.113***	0.113***	0.113***	0.113***	0.113***
	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)
Default in Previous Loans	0.0532***	0.0536***	0.0540***	0.0542***	0.0543***	0.0542***	0.0542***	0.0542***	0.0542***	0.0542***
	(0.0150)	(0.0150)	(0.0151)	(0.0151)	(0.0151)	(0.0151)	(0.0151)	(0.0151)	(0.0151)	(0.0151)
Job	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	28,430	28,430	28,430	28,430	28,430	28,430	28,430	28,430	28,430	28,430
Pseudo R^2	0.0139	0.0144	0.0146	0.0147	0.0147	0.0148	0.0149	0.0150	0.0150	0.0151
Wald Chi^2	488.79	497.59	503.58	507.63	509.33	512.34	515.84	519.04	522.16	525.53

This table presents the marginal effects of the average distance between the borrowers' 5-digit ZIP code neighborhood center and *k* nearest Maskan Bank branches on having relationship with Maskan Bank. The dependent variable is a binary variable that takes the value of one if the borrower has a relationship with Maskan Bank and zero otherwise. The variable of interest is *Distance to Maskan Bank Branches* (*MBD*). We control for *Male, Married, Bounced Check,* and *Default in Previous Loans*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 11: The Marginal Effects of Distance Ratio (DR) on Having Relationship with Maskan Bank

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
	k = 1	*k* = 2	*k* = 3	*k* = 4	*k* = 5	*k* = 6	*k* = 7	*k* = 8	*k* = 9	*k* = 10
DR	-0.0021**	-0.0042***	-0.0045***	-0.0049***	-0.0052***	-0.0055***	-0.0059***	-0.0065***	-0.0073***	-0.0079***
	(0.000820)	(0.00112)	(0.00133)	(0.00154)	(0.00169)	(0.00183)	(0.00194)	(0.00205)	(0.00217)	(0.00227)
Male	-0.000124	-0.000228	-0.000226	-0.000160	-0.000119	-0.000104	-0.0000954	-0.0000971	-0.000101	-0.000106
	(0.00599)	(0.00599)	(0.00599)	(0.00599)	(0.00599)	(0.00599)	(0.00599)	(0.00599)	(0.00599)	(0.00599)
Married	-0.0421***	-0.0418***	-0.0419***	-0.0420***	-0.0420***	-0.0420***	-0.0420***	-0.0420***	-0.0420***	-0.0420***
	(0.00742)	(0.00742)	(0.00742)	(0.00742)	(0.00742)	(0.00742)	(0.00742)	(0.00742)	(0.00742)	(0.00742)
Bounced Check	0.112***	0.113***	0.113***	0.113***	0.112***	0.112***	0.112***	0.112***	0.112***	0.112***
	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)	(0.0190)
Default in Previous Loans	0.0509***	0.0512***	0.0513***	0.0515***	0.0515***	0.0515***	0.0515***	0.0516***	0.0516***	0.0516***
	(0.0150)	(0.0150)	(0.0150)	(0.0150)	(0.0150)	(0.0150)	(0.0150)	(0.0150)	(0.0150)	(0.0150)
Job	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
N	28,430	28,430	28,430	28,430	28,430	28,430	28,430	28,430	28,430	28,430
Pseudo R^2	0.0120	0.0121	0.0121	0.0120	0.0120	0.0120	0.0120	0.0120	0.0121	0.0121
Wald Chi^2	434.26	441.76	439.25	437.41	437.07	436.56	436.75	437.54	438.67	439.49

This table presents the marginal effects of the distance ratio (*DR*) on having relationship with Maskan Bank. The dependent variable is a binary variable that takes the value of one if the borrower has a relationship with Maskan Bank and zero otherwise. The variable of interest is the distance ratio (*DR*). We control for *Male, Married, Bounced Check,* and *Default in Previous Loans*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 12: Average Marginal Effects of LPM, Probit, and Logit Regressions – Controlling for Borrowers' Wealth

	(1) LPM	(2) Probit	(3) Logit
Borrower's Relationship			
RelationshipDummy	-0.0437***	-0.0449***	-0.0450***
	(0.00456)	(0.00487)	(0.00500)
Borrower's Characteristics			
Male	0.00497	0.00609	0.00560
	(0.00611)	(0.00583)	(0.00604)
Married	0.0174***	0.0183***	0.0179***
	(0.00464)	(0.00482)	(0.00498)
Home Price	0.0518***	0.0534***	0.0558***
	(0.00952)	(0.00961)	(0.00962)
Borrower's Risk Measures			
Loan to Value	0.000575***	0.000559***	0.000545***
	(0.0000788)	(0.0000760)	(0.0000761)
Bounced Check	0.134***	0.115***	0.110***
	(0.0454)	(0.0333)	(0.0322)
Default in Previous Loans	0.452***	0.314***	0.295***
	(0.0123)	(0.00738)	(0.00696)
Effective Rate	-0.0527***	-0.0558***	-0.0545***
	(0.00284)	(0.00326)	(0.00322)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	26,581	26,560	26,560
R^2 (Pseudo R^2)	0.088	0.084	0.082

This table exhibits the average marginal effects resulted from estimation of Equation (1) using the LPM, probit, and logit estimation techniques when we include *Home Price* as a control variable. The sample is limited to Tehran due to data limiation. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is *RelationshipDummy*. We control for *Male, Married, Home Price, Loan to Value, Bounced Check, Default in Previous Loans,* and *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error terms are clustered at the Tehran municipal district level in all models. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 13: Results of LPM, Probit, and Logit – Endogeneity Concern: Sub-Sample of Borrowers with Prior Relationships

	(1) LPM	(2) Probit	(3) Logit
Borrower's Relationship			
RelationshipDuration (Year)	-0.0022***	-0.0020***	-0.00220***
	(0.0005)	(0.0004)	(0.000508)
Borrower's Characteristics			
Male	-0.0018	0.000003	-0.0008
	(0.0035)	(0.0034)	(0.0035)
Married	-0.0004	0.0002	-0.0002
	(0.0054)	(0.0055)	(0.0056)
Borrower's Risk Measures			
Loan to Value	0.0005***	0.000589***	0.0005***
	(0.0001)	(0.000108)	(0.0001)
Bounced Check	0.178***	0.150***	0.149***
	(0.0143)	(0.0106)	(0.0109)
Default in Previous Loans	0.442***	0.331***	0.317***
	(0.0091)	(0.00561)	(0.0051)
Effective Rate	-0.0675***	-0.0720***	-0.0707***
	(0.0029)	(0.00302)	(0.0030)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	49,821	49,821	49,821
R^2 (Pseudo R^2)	0.170	0.145	0.144
F (Wald Chi2)	169.27	5457.59	4960.78

This table exhibits the average marginal effects resulted from Equation (1) using the LPM, probit, and logit estimations. The sample is limited to the borrowers with prior relationships with Maskan Bank. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination, where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is the *RelationshipDuration*. We control for *Male, Married, Loan to Value, Bounced Check, Default in Previous Loans,* and *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error term is clustered at the city level in all models. Standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 14: Average Marginal Effects of LPM, Probit, and Logit Regressions for 6-Month and 12-Month Default Horizons

	6-Month			12-Month		
	(1)	(2)	(3)	(4)	(5)	(6)
	LPM	Probit	Logit	LPM	Probit	Logit
Borrower's Relationship						
RelationshipDummy	-0.0333***	-0.0302***	-0.0289***	-0.0488***	-0.0504***	-0.0485***
	(0.0052)	(0.0050)	(0.0049)	(0.0048)	(0.0053)	(0.0053)
Borrower's Characteristics						
Male	0.0104**	0.0051	0.0054	0.0050	0.0018	0.0039
	(0.0043)	(0.0041)	(0.0040)	(0.0046)	(0.0046)	(0.0043)
Married	0.0027	-0.0015	0.000003	0.0182**	0.0164**	0.0172**
	(0.0120)	(0.0104)	(0.0098)	(0.0090)	(0.0079)	(0.0076)
Borrower's Risk Measures						
Loan to Value	0.0003**	0.0002**	0.0003***	-0.0006***	-0.0007***	-0.0006***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Bounced Check	0.130***	0.152***	0.160***	0.131***	0.174***	0.193***
	(0.0144)	(0.0175)	(0.0196)	(0.0089)	(0.0112)	(0.0119)
Default in Previous Loans	0.234***	0.290***	0.299***	0.274***	0.365***	0.410***
	(0.0072)	(0.0077)	(0.0077)	(0.0064)	(0.0096)	(0.0121)
Effective Rate	-0.212***	-0.176***	-0.172***	-0.0280***	-0.0305***	-0.0306***
	(0.0058)	(0.0019)	(0.0017)	(0.00268)	(0.00288)	(0.0029)
Job	Yes	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes	Yes
N	31,025	31,025	31,025	47,460	47,454	47,454
R^2 (Pseudo R^2)	0.365	0.388	0.394	0.103	0.118	0.118
F (Wald Chi2)	211.23	6368.67	5515.91	140.87	3701.55	3232.14

This table exhibits the average marginal effects resulted from Equation (1) using the LPM, probit, and logit estimations. In the first three columns, the dependent variable is a binary variable that takes the value of one if the borrower defaults within the first 6 months from loan origination, and zero otherwise. In columns (4) to (6), the dependent variable takes the value of one if the borrower defaults within the first 12 months from loan origination. In all columns, the default is defined as failure to repay the installments for three consecutive months. The variable of interest is *RelationshipDummy.* We control for *Male, Married, Loan to Value, Bounced Check, Default in Previous Loans,* and *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error terms are clustered at the city level in all models. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 15: Results of LPM, Probit, and Logit – Alternative Measure of Relationship

	(1) LPM	(2) Probit	(3) Logit
Borrower's Relationship			
RelationshipDuration (Years)	-0.0032***	-0.0031***	-0.0033***
	(0.0003)	(0.0003)	(0.0003)
Borrower's Characteristics			
Male	0.0029	0.0036	0.0035
	(0.0027)	(0.0027)	(0.0027)
Married	0.0026	0.0031	0.0031
	(0.0057)	(0.0059)	(0.0059)
Borrower's Risk Measures			
Loan to Value	0.0004***	0.0004***	0.0004***
	(0.00008)	(0.00008)	(0.00008)
Bounced Check	0.169***	0.149***	0.151***
	(0.0113)	(0.0093)	(0.0093)
Default in Previous Loans	0.511***	0.397***	0.379***
	(0.0067)	(0.0044)	(0.0042)
Effective Rate	-0.0725***	-0.0770***	-0.0757***
	(0.0026)	(0.0027)	(0.0028)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	146,242	146,242	146,242
R^2 (Pseudo R^2)	0.197	0.163	0.162
F (Wald Chi2)	338.00	7837.95	6944.23

This table exhibits the average marginal effects resulted from Equation (1) using the LPM, probit, and logit estimations. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination, where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is the *RelationshipDuration*. We control for *Male, Married, Loan to Value, Bounced Check, Default in Previous Loans,* and *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error term is clustered at the city level in all models. Standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 16: Average Marginal Effects for Hard-Default Failure Event

	(1) LPM	(2) Probit	(3) Logit
Borrower's Relationship			
RelationshipDummy	-0.0169***	-0.0164***	-0.0173***
	(0.00242)	(0.00188)	(0.0018)
Borrower's Characteristics			
Male	-0.0057***	-0.0035**	-0.0048***
	(0.0015)	(0.0013)	(0.0014)
Married	-0.0023	-0.0016	-0.0020
	(0.0020)	(0.0020)	(0.0021)
Borrower's Risk Measures			
Loan to Value	0.00005	0.00005	0.00004
	(0.0000452)	(0.0000456)	(0.0000460)
Bounced Check	0.0845***	0.0515***	0.0475***
	(0.0088)	(0.0039)	(0.0039)
Default in Previous Loans	0.215***	0.113***	0.110***
	(0.0063)	(0.0028)	(0.0027)
Effective Rate	-0.0225***	-0.0258***	-0.0249***
	(0.0017)	(0.0011)	(0.0011)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	129,161	129,161	129,161
R^2 (Pseudo R^2)	0.1181	0.191	0.190
F (Wald Chi2)	48.80	4763.87	5143.32

This table exhibits the average marginal effects resulted from Equation (1) using the LPM, probit, and logit estimations. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination, where the default is defined as failure to repay the installments for six consecutive months. The variable of interest is *RelationshipDummy*. We control for *Male, Married, Loan to Value, Bounced Check, Default in Previous Loans*, and *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error terms are clustered at the city level in all models. The standard errors are calculated using the Delta method. Please refer to Appendix A1 for variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 17: The Marginal Effects of Prior Relationships – The Role of Religiosity

	(1) 200*m*	(2) 300*m*	(3) 400*m*	(4) 500*m*	(5) 600*m*
RelationshipDummy	-0.0322***	-0.0293***	-0.0289***	-0.0295***	-0.0325***
	(0.00399)	(0.00460)	(0.00487)	(0.00495)	(0.00549)
NM	-0.0001	0.00296	0.00129	-0.00006	-0.00103
	(0.00663)	(0.00357)	(0.00254)	(0.00222)	(0.00156)
RelationshipDummy × NM	-0.0143*	-0.0121***	-0.00769***	-0.00437**	-0.00145
	(0.00775)	(0.00360)	(0.00270)	(0.00197)	(0.00163)
Male	0.00451	0.00456	0.00459	0.00461	0.00458
	(0.00464)	(0.00463)	(0.00466)	(0.00466)	(0.00466)
Married	0.0196***	0.0195***	0.0196***	0.0196***	0.0196***
	(0.00555)	(0.00561)	(0.00561)	(0.00561)	(0.00558)
Loan to Value	0.0005***	0.0005***	0.0005***	0.0005***	0.0005***
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Bounced Check	0.174***	0.175***	0.175***	0.174***	0.174***
	(0.0253)	(0.0253)	(0.0253)	(0.0253)	(0.0252)
Default in Previous Loans	0.476***	0.476***	0.476***	0.476***	0.476***
	(0.0177)	(0.0177)	(0.0177)	(0.0178)	(0.0177)
Effective Rate	-0.0507***	-0.0507***	-0.0507***	-0.0508***	-0.0508***
	(0.00143)	(0.00144)	(0.00144)	(0.00148)	(0.00148)
Job	Yes	Yes	Yes	Yes	Yes
Education	Yes	Yes	Yes	Yes	Yes
N	27,918	27,918	27,918	27,918	27,918
R^2	0.0918	0.0918	0.0918	0.0918	0.0918
F	210.22	211.02	210.51	213.11	215.05

This table exhibits the average marginal effects resulted from Equation (8) using the linear probability model. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination, where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is the interaction term between *RelationshipDummy* and number of mosques inside circles with different radii (*NM*). Columns 1 to 5 represent the analyses for circles with radii of 200m to 600m. We control for *Married*, *Neighborhood Home Price*, *Loan to Value*, *Bounced Check*, *Default in Previous Loans*, *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. Standard errors are calculated using the Delta method. Please refer to Appendix A1 variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 18: The Marginal Effects of Prior Relationship – The Role of Gender

	(1) LPM	(2) Probit	(3) Logit
RelationshipDummy	-0.0308***	-0.0319***	-0.0315***
	(0.0080)	(0.0077)	(0.0085)
Male	0.0173***	0.0181***	0.0180**
	(0.0050)	(0.0050)	(0.0051)
RelationshipDummy × Male	-0.0229**	-0.0232**	-0.0241**
	(0.0091)	(0.0095)	(0.0099)
Married	0.0178***	0.0185***	0.0181***
	(0.00636)	(0.00640)	(0.00682)
Loan to Value	0.00058***	0.00056***	0.00055***
	(0.000158)	(0.000149)	(0.000153)
Bounced Check	0.134***	0.115***	0.110***
	(0.0395)	(0.0280)	(0.0278)
Default in Previous Loans	0.453***	0.315***	0.296***
	(0.0239)	(0.0150)	(0.0139)
Effective Rate	-0.0527***	-0.0558***	-0.0546***
	(0.00162)	(0.00186)	(0.00193)
Home Price	0.0514***	0.0531***	0.0555***
	(0.0193)	(0.0193)	(0.0197)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	26,574	26,553	26,553
R^2 (Pseudo R^2)	0.088	0.084	0.082
F (Wald Chi2)	95.43	2736.75	2568.84

This table exhibits the average marginal effects resulted from Equation (9) using the LPM, probit, and logit estimations. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination, where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is the interaction term between *RelationshipDummy* and *Male*. We control for *Married*, *Neighborhood Home Price*, *Loan to Value*, *Bounced Check*, *Default* in *Previous Loans*, *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. Standard errors are calculated using the Delta method. Please refer to Appendix A1 variable definitions. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Table 19: The Marginal Effects of Prior Relationship – The Role of Marital Status

	(1) LPM	(2) Probit	(3) Logit
RelationshipDummy	-0.0359**	-0.0372**	-0.0376**
	(0.0165)	(0.0157)	(0.0167)
Male	0.00891**	0.00979**	0.00955**
	(0.00430)	(0.00410)	(0.00437)
Married	0.0211**	0.0218**	0.0211**
	(0.0100)	(0.0101)	(0.0107)
RelationshipDummy × Married	-0.0095	-0.0094	-0.0090
	(0.0127)	(0.0126)	(0.0132)
Loan to Value	0.00057***	0.00056***	0.00055***
	(0.00016)	(0.00015)	(0.00016)
Bounced Check	0.134***	0.115***	0.110***
	(0.0396)	(0.0281)	(0.0279)
Default in Previous Loans	0.452***	0.314***	0.295***
	(0.0241)	(0.0151)	(0.0140)
Effective Rate	-0.0527***	-0.0558***	-0.0545***
	(0.00162)	(0.00186)	(0.00192)
Home Price	0.0519***	0.0535***	0.0559***
	(0.0194)	(0.0194)	(0.0197)
Job	Yes	Yes	Yes
Education	Yes	Yes	Yes
N	26,574	26,553	26,553
R^2 (Pseudo R^2)	0.088	0.084	0.082
F (Wald Chi2)	88.81	2548.80	2407.15

This table exhibits the average marginal effects resulted from Equation (10) using the LPM, probit, and logit estimations. The dependent variable is a binary variable that takes the value of one if the borrower defaults within the full observation period after loan origination, where the default is defined as failure to repay the installments for three consecutive months. The variable of interest is the interaction term between *RelationshipDummy* and *Married*. We control for *Married*, *Neighborhood Home Price*, *Loan to Value*, *Bounced Check*, *Default* in *Previous Loans*, *Effective Rate*. We also include *Job* and *Education* fixed effects. They are jointly significant; however, for brevity, we do not report their coefficients. The error term is clustered at the city level in all models. Standard errors are calculated using the Delta method. Please refer to Appendix A1 for the definition of the variables. ***, ** and * indicate significance at 1%, 5% and 10%, respectively.

Appendices

Appendix A1. Variable Descriptions

Variable	Description
3-Month Default	A dummy variable that takes the value of one if a borrower fails to make mortgage repayments for three consecutive months.
6-Month Default	A dummy variable that takes the value of one if a borrower fails to make mortgage repayments for six consecutive months.
RelationshipDummy	A dummy variable that takes the value of one if a borrower has a savings or current account prior to mortgage origination date.
RelationshipDuration	The number of years having a savings or current account prior to mortgage origination date.
Male	A dummy variable that takes the value of one if a borrower is male.
Married	A dummy variable that takes the value of one if a borrower is married.
Loan to Value	The ratio of a loan to the value of its collateral.
Bounced Check	A dummy variable that takes the value of one if a borrower has a history of bounced check with any bank in Iran.
Default in Previous Loans	A dummy variable that takes the value of one if a borrower has a default record in her history with any bank in Iran.
Effective Rate	The effective rate of a mortgage loan. It consists of the nominal interest rate of the loan and average market price of purchased mortgage paper.
Job	A categorical variable which specifies the borrower’s job category. The jobs in our data are categorized into major groups such as engineer, worker, manager, farmer, etc.
Education	A categorical variable which specifies the borrower’s education level. The levels of education are categorized into major groups such as elementary school, high school, bachelor’s degree, master’s degree, and the PhD.
Home Price	The average home price per square meter in the borrower’s neighborhood, calculated with the accuracy of the 5-digit Zip code. This data is reported monthly by the Ministry of Roads and Urban Development.

This table presents the description of the variables used in this study.

Appendix A2. Proportional Hazard Assumption Tests

In this appendix, we investigate whether the proportional hazards assumption holds (Hosmer et al., 2008). This assumption posits that the hazard ratios for the covariates remain constant over time, while the hazard rates are allowed to vary over time. Violation of this assumption in Cox regression analysis can result in biased estimates (Kuitunen et al., 2021). We utilize graphical methods for our investigation.

First, we follow graphical methods suggested in the literature for evaluating the proportionality of hazards for the categorical variables (Cleves et al., 2008). We compare the observed and predicted survival curves based on our Cox regression model. The observed survival curves represent the actual default behavior, while the predicted survival curves are derived from the Cox model. The results are depicted in Figure A2.1.

[Insert Figure A2.1 About Here]

The alignment between the observed and predicted survival curves for all variables is notably strong, indicating that our Cox regression model effectively captures the default patterns. However, it is worth noting that the curve for borrowers with a recorded bounced check displays a slight deviation from perfect alignment, although it remains within an acceptable range. Overall, the close correspondence between the observed and predicted survival curves supports the validity of the proportional hazards assumption and the hazard ratios have successfully passed the test of comparing observed and predicted survival probabilities.

Next, we plot an estimate of $-\ln\left(-\ln \hat{S}(t)\right)$ versus $\ln(t)$ for each level of categorical covariates in the equation, including *RelationshipDummy, Male, Married, Bounced Check*, and *Default in Previous Loans*, where $\hat{S}(t)$ is the Kaplan-Meier estimate of the survivor function.

Figure A2.2 exhibits the results for categorical variables related to borrowers' characteristics and risk measures including *RelationshipDummy*, *Male, Married, Bounced Check*, and *Default in Previous Loans*. Observing parallel curves in the log-log plots or survival curves indicates that the hazard ratio for the variable is constant over time, validating the proportional hazards assumption in the Cox regression model.

[Insert Figure A2.2 About Here]

A limitation of Figure A2.2 lies in its failure to address whether the effect of each variable remains constant when other variables are included in the model. Thus, further investigation is warranted to obtain separate graphs for each variable conditioned on other variables in the model. Figure A2.3 provides evidence that the effects of other included variables are controlled for, enhancing the robustness of the analysis. The consistency in the parallelism of the curves observed in Figure A2.3 re-affirms the reliability of this test. This strengthens our confidence in the proportional hazards assumption.

[Insert Figure A2.3 About Here]

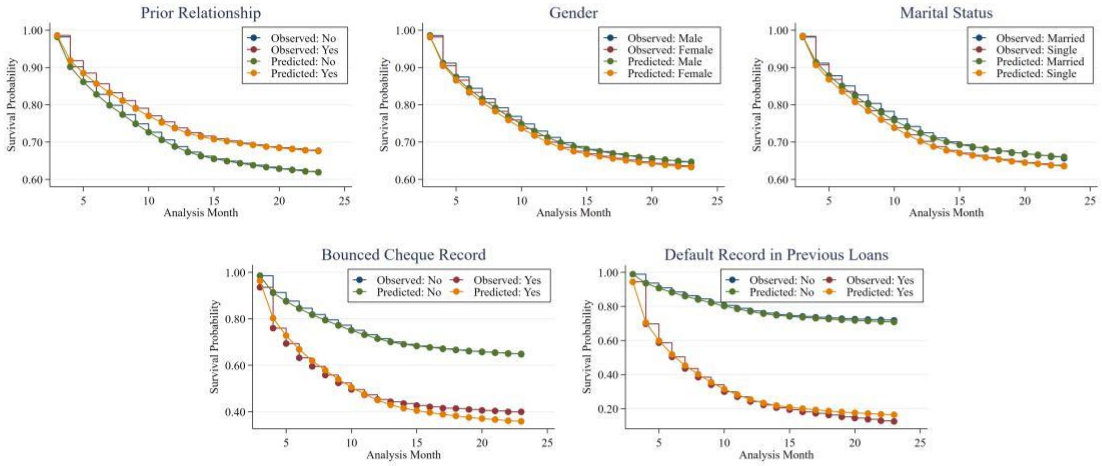
Additionally, we plot the scaled Schoenfeld residual (Schoenfeld, 1982) as an alternative test for the proportional hazards assumption, introduced by Grambsch and Therneau (1994). This approach essentially analyzes the presence of a non-zero slope in a generalized linear regression of the scaled Schoenfeld residuals plotted against time. It offers a reliable means of assessing whether hazard ratios remain constant over time, thereby confirming the assumptions inherent in the Cox regression model.

Figure A2.4 depicts the results. The smoothed residual curve demonstrates a consistent slope of zero over time, indicating that the hazard ratios remain constant. This provides further validation of the Cox regression model and reinforces our confidence in its ability to accurately model the relationship between covariates and survival times.

[Insert Figure A2.4 About Here]

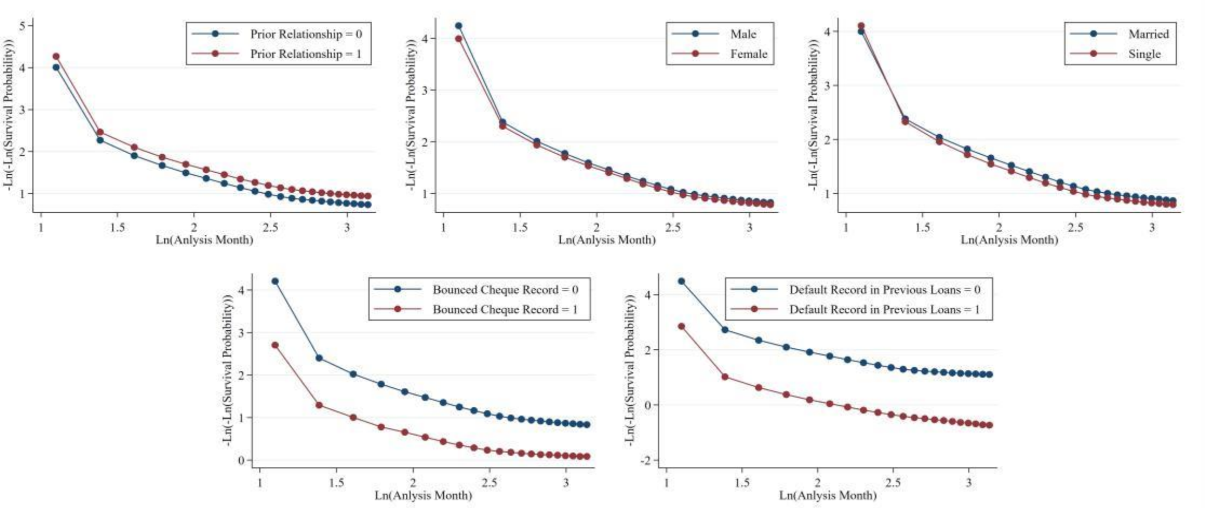
In conclusion, using a comprehensive array of graphical methods, we evaluate the consistency of hazard ratios over time. The results of our assessments indicate that the hazard ratios for all covariates remain constant over time. Therefore, we conclude that the proportional hazards assumption holds for the Cox regression model employed in our analysis. This suggests that the interpretation of hazard ratios remains valid, and the Cox regression results are reliable for predicting default behavior in our study.

Figure A2.1: Test of Proportional Hazards Assumptions for Categorical Variables – Comparison between the Observed and Predicted Survival Probabilities



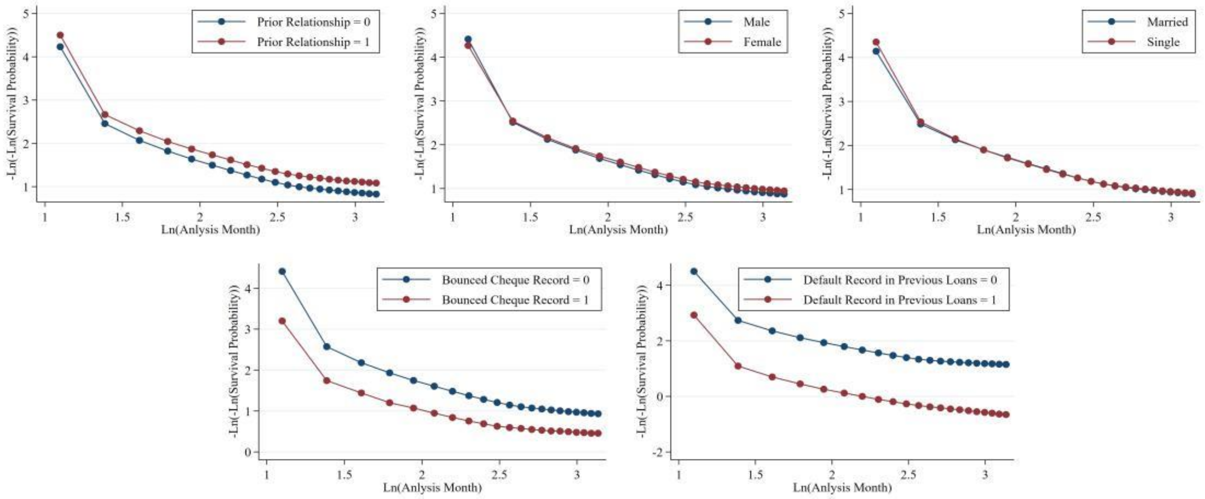
This figure presents a series of graphs that assess the proportional hazards assumption. They examine the similarity of observed and predicted survival probabilities for several categorical variables. The observed survival probability is depicted by the straight lines, whereas the predicted survival probability is represented by the curved lines. The y-axis represents the predicted and observed survival probabilities, while the x-axis shows the analysis month. Each graph focuses on a single categorical variable, indicated in the top of each graph.

Figure A2.2: Test of Proportional Hazards Assumptions for Categorical Variables – Log-Log Survival Probability



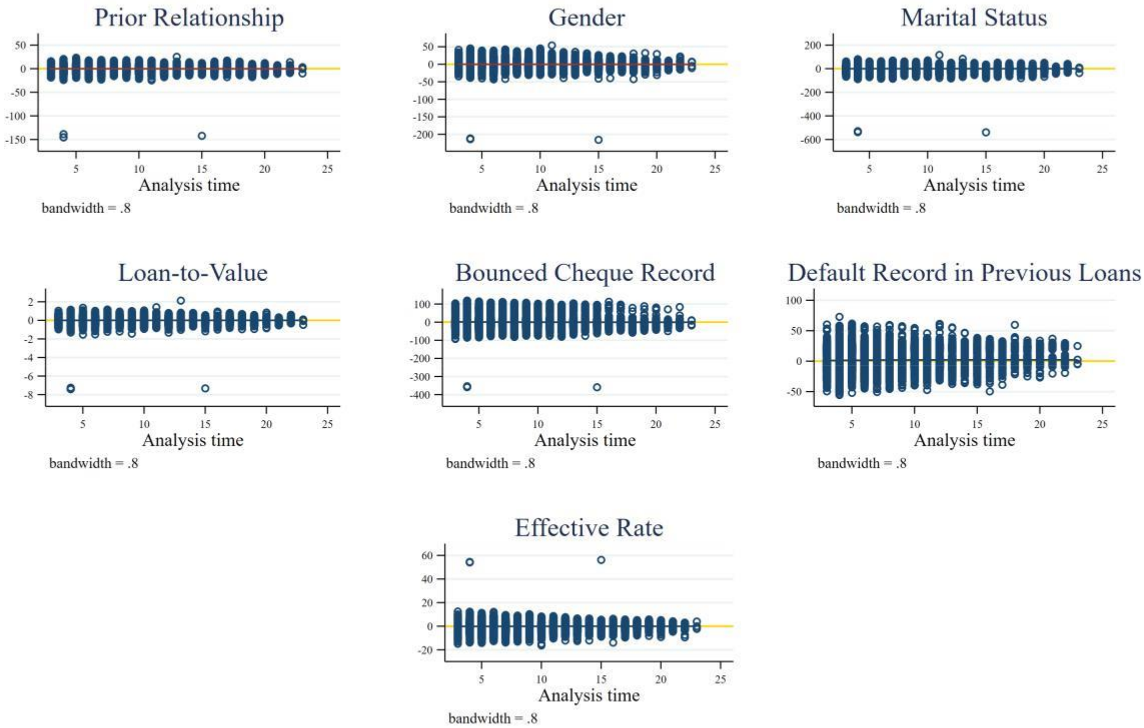
This figure presents a series of graphs that assess the proportional hazards assumption for several categorical variables in a Cox proportional hazards model for borrowers with prior relationships and those without prior relationships. The y-axis represents the negative log-transformed of the negative log of the survival probability (*ln*(-*ln*(-*S*(*t*)))). The x-axis shows the natural log of the analysis month (*ln*(*analysis month*)). Each graph focuses on a single categorical variable, indicated in the top right corner of each graph.

Figure A2.3: Test of Proportional Hazards Assumptions for Categorical Variables – Adjusted Log-Log Survival Probability



This figure presents a series of graphs that assess the proportional hazards assumption for several categorical variables in a Cox proportional hazards model. The y-axis represents the adjusted (*ln*(-*ln*(-*S*(*t*)))). This transformation is used to assess the proportional hazards assumption graphically, after adjusting for the effects of other variables included in the Cox model. The x-axis shows the natural log of the analysis month (*ln*(*analysis month*)). Each graph focuses on a single categorical variable, indicated in the top right corner of each panel.

Figure A2.4: Test of Proportional Hazards Assumptions – Graphical Scaled Schoenfeld Residual



This figure presents a series of graphs that assess the proportional hazards assumption for several variables in a Cox proportional hazards model. The y-axis represents the scaled Schoenfeld residuals and the x-axis shows the analysis time (months). Each graph focuses on a single variable, with the variable name in the title at the top.

Appendix A3. Machine Learning Methods - Performance Evaluation

To evaluate the performance of the models on the test set, we employ several metrics, with a primary focus on the Receiver Operating Characteristic (ROC) curve. The ROC curve provides a visual representation of the trade-off between the true positive rate (TPR) and the false positive rate (FPR) across varying probability thresholds for predicting defaults. A crucial summary statistic derived from the ROC curve is the Area Under the Curve (AUC), as highlighted by Bradley (1997). The AUC serves as an indicator of model effectiveness, with higher values denoting superior performance. Specifically, a higher AUC suggests that a higher TPR for any given FPR, and therefore, a more accurate classification of defaults.

The AUC, while a valuable metric, has limitations, particularly in datasets with a low prevalence of defaulters, where false positive rates (FPRs) are naturally low (Davis and Goadrich, 2006). This can make the AUC less informative in such contexts. To provide a more nuanced evaluation, we also compute the Precision and Recall for each model. Precision, defined as $P(y=1|\hat{y}=1)$, measures the proportion of positive predictions that are true defaulters. Recall, defined as $P(\hat{y}=1|y=1)$, represents the proportion of actual defaulters that the model correctly identifies. We use these metrics to construct Precision-Recall curves, which plot Precision against Recall over varying probability thresholds. To summarize these curves, we calculate the average Precision score, offering a weighted mean that balances Precision and Recall.

Figure A3.1 illustrates the performance of four models – logit, XGBoost, random forest, and neural network – on the test dataset, showcasing their ROC and Precision-Recall curves. The ROC AUC scores indicate that XGBoost achieve the highest performance with an AUC of 0.7603, followed by the neural network with 0.7512, logit with 0.7434, and random forest with 0.7316. This suggests that XGBoost is the most effective at distinguishing between positive and negative classes, while random forest performs the least well in this regard.

[Insert Figure A3.1 About Here]

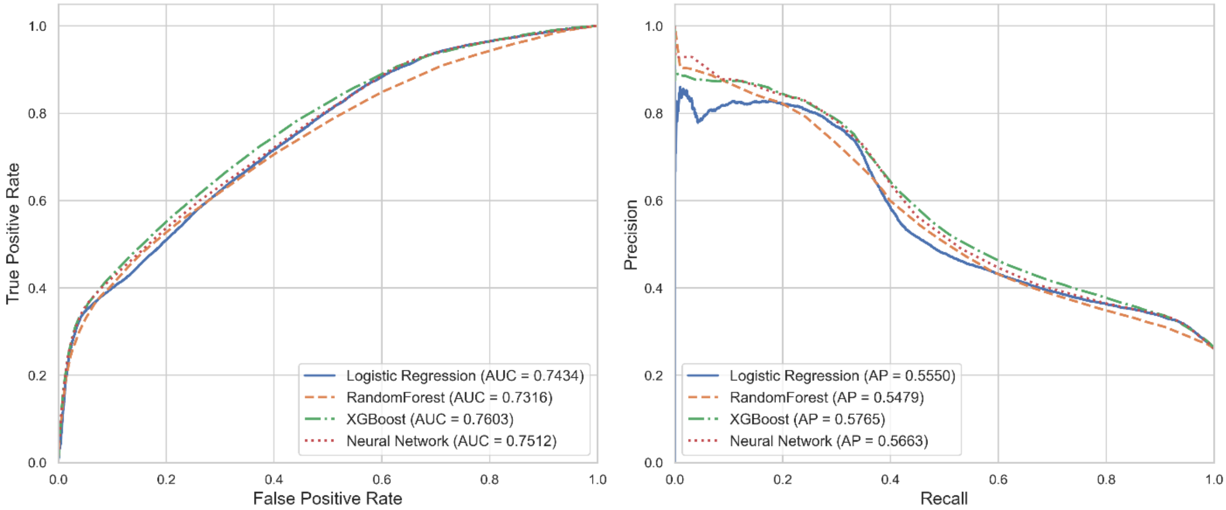
In terms of Precision-Recall curves (Panel B), XGBoost also leads with an average precision (AP) score of 0.5765, slightly ahead of the neural network with an AP of 0.5663. Logit and random forest show lower AP scores of 0.5550 and 0.5479, respectively. These results indicate that XGBoost not only has the best capability for separating classes but also maintains a higher precision when the model predicts positive outcomes. The neural network, while close to XGBoost in AP, falls short in ROC AUC compared to XGBoost. Overall, XGBoost demonstrates superior performance in both metrics, highlighting its robustness in handling the classification task.

We also evaluate the models using the Brier Score and R-squared (R^2). The Brier Score measures the average squared error of predictions, with a lower score indicating better model accuracy. Meanwhile, the R^2 statistic represents the proportion of variance in the dependent variable that is explained by the model, calculated as one minus the ratio of the sum of squared residuals from the model to the sum of squared residuals from a model predicting the mean of the target variable. A higher R^2 value signifies a greater portion of the variability in the data is accounted for by the model.

Table A3.1 presents the performance metrics – including ROC AUC, Average Precision Score (AP), Brier Score, and R^2 – for the logit, random forest, XGBoost, and neural network models. The comparison of performance metrics reveals that XGBoost and neural network models exhibit better predictive accuracy and model fit compared to logit and random forest. As shown in column (3), XGBoost achieves the lowest Brier Score of 0.1518, and as indicated in column (4), it also has the highest R-squared value of 0.2121, demonstrating its superior performance in minimizing prediction errors and explaining the variance in the data. The neural network model similarly performs well, with a Brier Score of 0.1526 and an R-squared of 0.2079. In contrast, both logit and random forest exhibit higher Brier Scores and lower R-squared values, with logit at 0.1563 and 0.1890, and random forest at 0.1581 and 0.1795. These results suggest that XGBoost and neural network are more effective at leveraging the training data to produce more accurate predictions.

[Insert Table A3.1 About Here]

Figure A3.1: Receiver Operating Characteristic and Precision-Recall Curves



A) Receiver Operating Characteristic

B) Precision-Recall

This figure shows the Receiver Operating Characteristic (Panel A) and Precision-Recall curves (Panel B) as performance metrics for logit, random forest, XGBoost, and neural network methods.

Table A3.1: Performance Metrics of Statistical Technologies

Model	ROC AUC	Average Precision Score	Brier Score	R^2
Logit	0.7434	0.5550	0.1563	0.1890
Random Forest	0.7316	0.5479	0.1581	0.1795
XGBoost	0.7603	0.5765	0.1518	0.2121
Neural Networks	0.7507	0.5663	0.1526	0.2080

Table displays the performance metrics for logit, random forest, XGBoost, and neural network models. The metrics include ROC AUC, Average Precision Score, Brier Score, and R-squared (R^2). For ROC AUC, Average Precision Score, and R^2 higher values indicate better predictive accuracy and model-fit. Conversely, for the Brier Score, lower values reflect better accuracy, as it measures the average squared prediction error with lower scores indicating fewer prediction errors.

Appendix A4. Summary Statistics of Neighborhoods' Home Price Per Square Meter at Tehran's Municipal Districts

Tehran's Municipal District	Mean	Min	Max	SD	N
1	733.8	262.50	1,427.52	301.33	1,583
2	840.4	289.97	1,400.00	251.24	2,141
3	911.3	273.58	1,431.38	258.80	708
4	662.9	256.52	1,432.06	199.39	2,069
5	750.5	267.83	1,425.00	196.73	3,219
6	756.3	260.10	1,432.67	239.70	827
7	600.1	277.27	1,385.10	186.21	1,245
8	657.0	257.20	1,425.00	172.69	1,516
9	500.3	270.32	1,159.09	136.20	1,017
10	476.9	254.58	1,200.00	114.08	2,465
11	492.2	265.89	1,365.78	155.04	1,451
12	439.8	255.00	1,327.83	155.07	867
13	555.9	270.00	1,352.04	153.52	1,193
14	512.8	256.70	956.10	117.19	1,392
15	435.3	255.87	1,050.00	129.11	1,425
16	397.4	257.13	975.14	98.31	356
17	392.9	256.95	856.48	89.32	1,077
18	364.5	254.96	1,249.84	93.32	913
19	460.7	260.53	1,430.00	238.29	275
20	449.7	260.00	1,399.80	164.88	113
21	525.7	261.51	864.64	95.52	654
22	657.4	271.88	1,292.54	191.08	530

This table provides a summary statistics of home prices per square meter in $ across different district zones in Tehran, Iran. The statistics include the mean, minimum, maximum, standard deviation (SD), and the number of observations (N) for each zone. The city of Tehran is divided into 22 municipal districts, each with its own administrative center.

Appendix A5. Marginal Effects of Prior Relationships on Loan Default at Different Home Prices

At Home Price ($ per m^2)	Relationship Marginal Effect	Delta-Method SE	Z	P-Value	[95% conf. interval]	
250	-0.0285	0.00860	-3.31	0.000	-0.0453	-0.0116
300	-0.0306	0.00797	-3.84	0.000	-0.0463	-0.0150
350	-0.0328	0.00734	-4.47	0.000	-0.0472	-0.0185
400	-0.0351	0.00672	-5.22	0.000	-0.0483	-0.0219
450	-0.0374	0.00611	-6.13	0.000	-0.0494	-0.0254
500	-0.0397	0.00551	-7.21	0.000	-0.0506	-0.0289
550	-0.0421	0.00496	-8.49	0.000	-0.0519	-0.0324
600	-0.0446	0.00447	-9.97	0.000	-0.0533	-0.0358
650	-0.0471	0.00408	-11.53	0.000	-0.0551	-0.0391
700	-0.0496	0.00383	-12.96	0.000	-0.0571	-0.0421
750	-0.0522	0.00376	-13.89	0.000	-0.0595	-0.0448
800	-0.0548	0.00389	-14.1	0.000	-0.0624	-0.0472
850	-0.0575	0.00422	-13.64	0.000	-0.0657	-0.0492
900	-0.0602	0.00471	-12.78	0.000	-0.0694	-0.0510
950	-0.0630	0.00533	-11.81	0.000	-0.0734	-0.0525
1,000	-0.0658	0.00606	-10.86	0.000	-0.0777	-0.0539
1,050	-0.0686	0.00686	-10.01	0.000	-0.0821	-0.0552
1,100	-0.0715	0.00771	-9.28	0.000	-0.0867	-0.0564
1,150	-0.0745	0.00862	-8.64	0.000	-0.0914	-0.0576
1,200	-0.0775	0.00956	-8.1	0.000	-0.0962	-0.0587
1,250	-0.0805	0.01054	-7.64	0.000	-0.1012	-0.0598
1,300	-0.0836	0.01155	-7.24	0.000	-0.1062	-0.0609
1,350	-0.0867	0.01259	-6.89	0.000	-0.1114	-0.0620
1,400	-0.0899	0.01365	-6.58	0.000	-0.1166	-0.0631
1,450	-0.0931	0.01473	-6.32	0.000	-0.1219	-0.0642

This table presents the marginal effects of having prior relationships with the bank on the likelihood of default at different neighborhoods home prices per square meter in $. The results are obtained from estimating the Equation (4) using logit regression. The table includes the marginal effect, Delta-method standard error (SE), Z-value, P-value, and the 95% confidence interval for the marginal effect.

Appendix A6. Summary Statistics of the Number of Mosques within Circles of Different Radii around Borrowers' Neighborhood Centers

	Mean	SD	Min	Max	N
100m Circle	0.05	0.25	0.00	4.00	28,430
200m Circle	0.20	0.48	0.00	6.00	28,430
300m Circle	0.48	0.81	0.00	8.00	28,430
400m Circle	0.81	1.13	0.00	12.00	28,430
500m Circle	1.28	1.58	0.00	22.00	28,430
600m Circle	1.80	2.02	0.00	28.00	28,430

The Centre for Responsible Banking & Finance
CRBF Working Paper Series
Department of Finance, University of St Andrews
The Gateway, North Haugh,
St Andrews, Fife,
KY16 9RJ.
Scotland, United Kingdom
https://crbf.wp.st-andrews.ac.uk/

Recent CRBF Working papers published in this Series

First Quarter | 2025

25-003 **Alfred Duncan, João Pedro De Camargo Mainente, and Charles Nolan:** Monetary and Financial Policy with Privately Optimal Risk Taking.

25-002 **Mehmet Selman Çolak, Yavuz Kılıç, Yusuf Yıldırım, and Muhammed Hasan Yılmaz:** The Effect of Personal Credit Card Limits on Card Use: Micro Data Evidence from an Emerging Market.

25-001 **Hans Degryse, Olivier De Jonghe, Leonardo Gambacorta, and Cédric Huylebroek:** Bank Specialization and Corporate Innovation.

Fourth Quarter | 2024

24-030 **Dimitris K. Chronopoulos, John O.S. Wilson, and Muhammed H. Yilmaz:** State Ownership and Bank Lending: Evidence from Turkish Banking.

24-029 **Anna L. Sobiech, Lechedzani Kgari, Linh Nguyen, and John O.S. Wilson:** Implicit Guarantees and Cross-Border Bank Stability: Evidence from Geographic Ringfencing.

24-028 **Thomas Krause, Eleonora Sfrappini, Lena Tonzer, and Cristina Zgherea:** How Do EU Banks' Funding Costs Respond to the CRD IV? An Assessment Based on the Banking Union Directives Database.

24-027 **Kulnicha Meechaiyo:** Cash is King or Trash? The Review of Political Uncertainty and Corporate Behaviour.

24-026 **Amanda Rae Heitz and John Kandrac:** Heterogeneous Competition in Banking: Evidence from Multiple Loan and Deposit Products.

University of St Andrews
Scotland's first university
600 YEARS
1413 – 2013