Pontifícia Universidade Católica
do Rio de Janeiro

Karolina Staniczek Andrade

Advantageous selection in the payroll loan market

Dissertação de Mestrado

Thesis presented to the Programa de Pós–graduação em Economia, do Departamento de Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia.

Advisor: Prof. Leonardo Rezende

Rio de Janeiro
Setembro 2023

Pontifícia Universidade Católica
do Rio de Janeiro

Karolina Staniczek Andrade

Advantageous selection in the payroll loan market

Thesis presented to the Programa de Pós–graduação em Economia da PUC-Rio in partial fulfillment of the requirements for the degree of Mestre em Economia. Approved by the Examination Committee:

Prof. Leonardo Rezende
Advisor
Departamento de Economia – PUC-Rio

Prof. Márcio Nakane
FEA-USP

Profa. Nathalie Gimenes
PUC-Rio

Rio de Janeiro, Setembro 13th, 2023

Karolina Staniczek Andrade

B.A. in Economics, Universidade de São Paulo (USP), 2019

Bibliographic data

Staniczek Andrade, Karolina

Advantageous selection in the payroll loan market / Karolina Staniczek Andrade; advisor: Leonardo Rezende. – 2023.

52 f: il. color. ; 30 cm

Dissertação (mestrado) - Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Economia, 2023.

Inclui bibliografia

1. Economia – Teses. 2. Informação assimétrica. 3. Empréstimos. 4. Crédito. 5. Mercado bancário. I. Rezende, Leonardo. II. Pontifícia Universidade Católica do Rio de Janeiro. Departamento de Economia. III. Título.

CDD: 004

Acknowledgments

I would like to thank my advisor Prof. Leonardo Rezende, whose insightful comments and contributions were essential to this research. In addition, I thank committee members, Márcio Nakane and Nathalie Gimenes, for their participation and valuable comments.

I am profoundly grateful to my parents, Michalina and Rogerio, for their unwavering support, attentive care, and steadfast belief in me, which made everything possible. I extend my sincere thanks to my dear sisters, Greta and Sofia, for willingly having my back throughout these years.

I am especially thankful to João, for his endless patience, love and companionship, and for holding my hand through the highs and lows of this journey. Finally, I would like to thank my friends who have been a perpetual source of joy and encouragement to me.

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001.

Abstract

Staniczek Andrade, Karolina; Rezende, Leonardo (Advisor). **Advantageous selection in the payroll loan market**. Rio de Janeiro, 2023. 52p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

This paper investigates the nature of information asymmetries in the Brazilian payroll loan market from 2013 to 2021. We develop a demand model that accounts for the effect of consumers' private information on borrowing decisions. The novelty of the model is its ability to extract information on unobservable characteristics using publicly available firm-level data. Empirically, we use the variation of market shares and default rates within banks to estimate a utility parameter that represents the sign of selection in the market. Our analysis reveals empirical evidence of advantageous selection within the market, indicating that safer borrowers are more inclined to apply for loans. Additionally, we expand the model to incorporate a distinct parameter for Caixa Econômica, a state-owned bank that exhibited different behavior compared to other financial institutions during the same period. Our analysis reveals a significantly lower selection parameter in magnitude for Caixa.

Keywords

Asymmetric information; Loans; Banking; Credit market; Default;.

Resumo

Staniczek Andrade, Karolina; Rezende, Leonardo. **Seleção vantajosa no mercado de empréstimos consignados**. Rio de Janeiro, 2023. 52p. Dissertação de Mestrado – Departamento de Economia, Pontifícia Universidade Católica do Rio de Janeiro.

Este artigo investiga a natureza das assimetrias de informação no mercado brasileiro de empréstimos consignados de 2013 a 2021. Desenvolvemos um modelo de demanda que leva em consideração o efeito da informação privada dos consumidores nas decisões de empréstimo. A novidade do modelo é sua capacidade de extrair informações sobre características não observáveis usando dados públicos a nível dos bancos. Empiricamente, utilizamos a variação das participações de mercado e das taxas de inadimplência dos bancos para estimar um parâmetro de utilidade que representa o sinal de seleção no mercado. Nossa análise revela evidências empíricas de seleção vantajosa dentro do mercado, indicando que os tomadores de empréstimos mais seguros estão mais inclinados a solicitar empréstimos. Além disso, expandimos o modelo para incorporar um parâmetro distinto para a Caixa Econômica, um banco estatal que exibiu comportamento diferente em comparação com outras instituições financeiras durante o mesmo período. Nossa análise revela um parâmetro de seleção significativamente menor em magnitude para a Caixa.

Palavras-chave

Informação assimétrica; Empréstimos; Crédito; Mercado bancário.

Table of contents

List of figures

List of tables

List of Abreviations

ESTBAN – Estatística Bancária

IF DATA – Dados selecionados de instituições financeiras

CNPJ — Cadastro Nacional da Pessoa Jurídica

OLS — Ordinary Least Squares

IV – Instrumental Variables

1
Introduction

Since the seminal paper of Stiglitz and Weiss (1981), the impact of informational asymmetries in credit markets has been widely studied. However, a theoretical debate about the sign of selection at play persists: some argue that riskier consumers are more prone to take out loans (adverse selection), whereas other show that safer consumers are attracted to the pool of borrowers (advantageous selection). These contrasting perspectives have different implications for market operations. Adverse selection can lead to credit rationing (Stiglitz and Weiss, 1981), where lenders limit credit availability when unable to distinguish between high and low-risk borrowers. On the other hand, advantageous selection can result in overlending (Meza and Webb, 1987), a case of excessive credit supply. In both cases, suboptimal capital allocation potentially arises, since the market fails to efficiently provide credit to individuals with a high marginal rate of return (Banerjee and Duflo, 2014) or deny credit to those who do not need it.

Empirical studies have similarly produced mixed results regarding the presence of selection for various markets and contexts. Understanding the nature of selection is crucial for comprehending the workings and underlying challenges of the market. Furthermore, there is evidence that selection forces interact with market conduct, creating dynamics where frictions can potentially offset each other (Crawford et al., 2018).

The objective of this paper is to identify the sign of selection present in the payroll loan market in Brazil. We develop a model that accounts for the interplay of demand for loans and default and allows us to estimate the presence of information asymmetries with publicly available aggregate data. Due to the relevance of private information in the borrower's decision, we adapt and expand the demand estimation framework, the multinomial logit proposed by McFadden (1973), to include the effect of consumers' ex-ante risk. Hence, the novelty of the model is its ability to extract information on consumers' unobservable characteristics from firm-level data. Moreover, our methodology provides an alternative approach to incorporate consumer heterogeneity into the demand estimation problem, a concept initially introduced in this line of models by Berry et al. (1995).

Our estimation is conducted in two steps. In the first step, we estimate a multinomial logit to identify utility parameters of consumers that choose to make loans in the market. The underlying assumption is that information asymmetry affects the choice to take out a loan but not the decision regarding which institution to borrow from. The estimated mean utilities for each bank in every quarter, reflecting consumer preferences for these institutions, are used to compute the predicted market shares. Based on the estimates, in the second step, we estimate a system of equations composed by the total demand for loans and the banks' default rates. These equations depend on the distribution of the risk types in the population, which reflects the probability of an individual failing to repay their loan. Empirically, we observe how variation in credit portfolio affects the default rate of the banks. As a result, we can infer information about the risk profile of marginal consumers who join the pool of borrowers when banks expand their credit lending.

The Brazilian market provides a compelling context for this study. Brazil presents high levels of bank spreads that persist despite the development in the financial sector. The average bank spread[1] in Brazil for all credit operations was 27% in 2022, while the average for middle income countries was 5.6% (2022 values) and for low income countries, 9.4% (2017 values)[2]. The credit market in Brazil also presents a significant degree of concentration: the top five banks — Caixa, BB, Itaú, Santander, and Bradesco — hold 75% of the personal credit portfolio, a similar trend observed in the deposit market.

This concentration, coupled with elevated interest rates, has raised concerns in public discourse, often attributed to uncompetitive practices. However, empirical evidence supporting this claim remains scarce. Banks assert that the situation is primarily driven by onerous costs such as taxes, high delinquency rates and challenges in guarantee recovery. Additionally, in a country with high levels of labor market informality, accurately screening loan applicants is potentially burdensome. Our research aims to investigate one potential source of market problems: information asymmetries.

We focus on personal credit, where banking spreads are higher. This stems from the fact that loans for individuals are usually more difficult to screen and get guarantees for, which is precisely why asymmetries tend to influence more. We look at payroll loans, the second highest portfolio among households in Brazil. These loans present reduced risk and monitoring costs for banks because repayment is automatically deducted from the borrower's payroll or social security benefits. Payroll loans are exclusively available to pensioners and formal workers[3], specifically those whose employers have contracts with banks.

[1]World Bank estimates from https://data.worldbank.org/indicator/FR.INR.LNDP'

[2]Bank spread for loans for individuals averaged 24% in the period from 2012 to 2021. If we exclude targeted loans, which are subsidized by the government, the average of bank spreads in the period is 41%

[3]This type of credit also presents predefined interest rate limits for pensioners and public employees

Therefore, we analyze a subset of borrowers with relatively stable income sources. While default is not an option that borrowers can intentionally choose, certain situations may hinder repayment, such as when a borrower leaves their current employer[4]. Payroll credit, despite being considered the safest option, still carries a significant level of risk for banks: approximately 0.77%[5] of the total portfolio is more than 90 days overdue[6].

Our model points to the presence of advantageous selection in this line of credit in Brazil. The estimated parameters indicate that the utility of taking out a loan is higher for individuals with a lower probability of default. Although we measure the effect by the demand side as a utility parameter, this result is driven by both consumer preferences and the monitoring and lending approval practices of banks. When a bank is highly selective in its borrowing approvals, the pool of borrowers tends to exhibit a lower aggregate risk of default. Similarly, when safer consumers, especially those with stable incomes, show a preference for borrowing, particularly in the case of loans with automatic repayment mechanisms like payroll loans, this also contributes to a reduced aggregate default risk. Therefore, we cannot disentangle both effects from the available aggregate data.

The second goal of this research is to examine differences in information asymmetries between private and state-owned banks. In the period from 2011 to 2014, Brazilian state-owned banks, mainly Caixa Econômica Federal and Banco do Brasil, were utilized by the government to stimulate credit and lower interest rates, in an attempt to encourage private banks to follow[7]. During the study period, Caixa launched the "Caixa Melhor Crédito" ("Better Credit") program, which focused on personal credit and increased payroll loan portfolio by 35% from December 2012 to December 2013[8]. The credit expansion was followed by an increase in their delinquency rate. Furthermore, even after the program was discontinued in 2014, default rates remained higher than the average observed in the sample of banks. This suggests Caixa's willingness to accommodate consumers with relatively higher risk profiles.

In this paper, we flexibilize the methodology proposed, which initially

[4]Additionally, in the case of borrower's death, there may be instances where loans were taken out in the name of the pensioner by family members, benefiting from the lower interest rates offered to pensioners. These loans may not be repaid following the borrower's demise. Furthermore, if an individual obtains loans from multiple banks and the cumulative interest payments exceed 30% of their payroll, the automatic payment will not be made and the loan may be classified as overdue

[5]Values from the fourth quarter of 2021

[6]Although this percentage is lower compared to unsecured credit (2.7%) or vehicle loans (1.4%), the volume of payroll credit is 3 times and 2.2 times larger, respectively.

[7]G1 Globo (2012b), G1 Globo (2012a), G1 Globo (2013)

[8]In the same period, Banco do Brasil, the second biggest bank, launched the program "Bom para todos", also to boost credit, but focused on micro and small firms, vehicles)

accounts for a homogeneous selection effect for all firms in the market, by allowing this effect to vary for Caixa Econômica. The adjustment is motivated by the fact that Caixa exhibited distinct credit expansion and pricing behavior from other banks in the market in the period studied. This implies that Caixa may be differently affected by informational asymmetries, possibly due to either varying consumer preferences among different risk profiles that favor Caixa, or a distinct loan approval policy employed by the bank to determine loan recipients.

There are some reasons to believe that state-owned banks are differently affected by informational asymmetries compared to private banks. Public banks[9], being less profit-driven, may have the flexibility to serve consumers who might not qualify for credit from private banks. This role is typically seen as a means of promoting financial inclusion and benefiting a broader segment of households. However, the impact of this approach on default rates and the banks' costs hinges on the type of selection prevalent in the market. If the state-owned banks operate under adverse selection, they might be accommodating a pool of borrowers with a higher probability of default.

Our results show that Caixa Econômica presents a much lower parameter for selection, indicating that it operates taking in more risky consumers than the rest of the banks in this market.

The present study faces some limitations. Testing for selection involves measuring private information held by consumers, which is precisely the information that banks struggle to obtain in the first place. While the literature uses loan-level datasets to measure the correlation between ex-ante interest rates and loan terms with ex-post default rates, after controlling for observable factors (Chiappori and Salanie, 2000), such datasets are challenging to access. In our context, we only have access to aggregate data that carry approximations that might lead to measurement errors. Moreover, studies in this area distinguish between selection (hidden information) and moral hazard (hidden action), as they involve different behavioral mechanisms. In this paper, we are unable to fully disentangle asymmetric information from moral hazard. Nevertheless, the type of loans chosen minimizes the effect of hidden action once default is not an option for the borrower.

The remainder of the paper is organized as follows: Chapter 2 presents a literature review. Chapter 3 introduces the data and statistics. Chapter 4 and 5 describe the model and its econometric specification. We present the estimates in Chapter 6. Chapter 7 concludes.

[9]As it is done in Coelho et al. (2013), in this work, the term "public banks" is used as a synonym for state-owned banks and not for publicly held banks

2
Related literature

Information asymmetries were first identified as a source of market distortion in the seminal paper on the "market for lemons" (Akerlof (1970)). This concept was later applied to the credit market by Stiglitz and Weiss (1981) to explain how selection affects the supply of credit.

In their model, banks assume that as interest rates increase, the quality of borrowers in the pool deteriorates due to adverse selection. To mitigate this effect, banks set loan interest rates below the efficient outcome market equilibrium, but also offer credit in smaller quantities. Consequently, the total credit available falls below the socially efficient level, leading to credit rationing. Banks are unable to distinguish between low and high-risk borrowers, so they opt to lend to a smaller pool of borrowers with lower chances of being high-risk.

However, Meza and Webb (1987) challenge this view and argue that asymmetric information can lead to the opposite outcome. Good payers not only differ in risk but also in expected return. If borrowers who are more willing to pay higher rates are the ones with higher expected revenue from their projects, the overall quality of borrowers increases as interest rates rise. In this scenario, choosing a lower interest rate could result in excessive provision of credit, known as overlending.

Since Stiglitz and Weiss (1981) and the theoretical advances that followed, numerous studies have aimed to empirically document the presence of asymmetric information in real-life contracts. However, the findings regarding the existence and extent of selection effects vary considerably across markets, contract modalities, and institutional contexts. Our research closely aligns with this body of literature, as we seek to test for existence of asymmetric information in insurance and credit markets.

For insurance markets, there is evidence of adverse selection for: annuity market (Finkelstein and Poterba, 2004), life insurance (He, 2009), automobile insurance in Israel (Cohen, 2005), crop insurance (Gunnsteinsson, 2020), whereas other show no evidence of informational asymmetries, such as automobile insurance in France (Chiappori and Salanie, 2000) and health insurance (Cardon and Hendel, 2001). Advantageous selection appear in health insurance for seniors (Fang et al., 2008) and U.S. reverse mortgage market (Davidoff and Welke, 2004).

For credit markets, empirical evidence of adverse selection is documented in the following markets: credit cards (Ausubel, 1999), home mortgage and automobile loans (Edelberg, 2004), home equity credit (Agarwal et al., 2006), subprime automobile loans (Adams et al., 2009), and personal loans (Karlan and Zinman, 2009).

In a more relatable setting to ours, Kim (2017) presents evidence of advantageous selection in personal loans, by exploring a loan-level dataset of a big bank in South Korea. He finds evidence that interest rates and default are negatively correlated, indicating that borrowers with lower risk profiles tend to self-select into loan contracts characterized by higher interest rates and a lower level of collateral. He argues that the mechanism that drives the selection is the combination of heterogenous consumers with strong preferences for consumption smoothing and an uncompetitive market, which allows banks to exert their market power in order to charge higher interest rates.

In another work that addresses the same market, Kim and Yoon (2022) explore a dataset from a major credit rating company, which provides information for all banks in the market and the consumers' relationships across multiple banks. They find evidence of adverse selection in South Korea. Despite the contrasting views for the same market, the pool of borrowers for each type of bank directly relates to the selection outcome. For instance, Kim (2017) analyzes the set of borrowers from one specific institution, which might operate differently from the rest of the market. Moreover, Finkelstein and McGarry (2006) argue that different types of consumers coexist in a market, generating opposing forces that drive selection. For example, higher-risk types and more risk-averse consumers both purchase higher coverage plans, suggesting an absence of correlation between default and coverage. Therefore, the types of consumers significantly influence the selection forces observed in the market.

The methodologies for identifying selection include large-scale randomized experiments, exploring institutional changes, and a tractable test developed by Chiappori and Salanie (2000). The latter has been frequently applied and its main idea is to look for correlation between ex-ante prices or interest rates accepted by the consumer and ex-post outcomes, such as defaults or insurance costs. All methods depend on the availability of loan-level data, usually confidential, which explain the relative scarcity of such studies, specially in credit markets. From the methodological perspective, our main contribution to this strand of the literature is to build a model which allows us to identify such asymmetry based on aggregate data.

Additionally, our paper builds on models such as McFadden (1973) and Berry et al. (1995), later applied to banking in Dick (2008) in the US and Nakane et al. (2006) in Brazil. We adapt the demand estimation framework from these models in order to measure the impact of unobservable characteristics that influence consumers' decisions to take out loans.

Secondly, we attempt to provide insights on the differences between public and private-owned banks in terms of information asymmetries. The role of state-owned banks has long been debated in the literature. Some argue that they exist because of market failures and they help finance projects that are socially valuable, but would not be undertaken by private intermediaries. This, in turn, would enhance the provision of financial services to less privileged consumers or areas, which would be justified as a development goal. On the other hand, some contend that state-owned banks might be less efficient and may not effectively achieve their intended objectives.

In a market with private information, state-owned banks might be less cautious about the type of borrower they are willing to offer loans and they may attract a different kind of consumer. As highlighted by Dell'ariccia et al. (1999), informational asymmetries can also impact the supply of borrowers. If riskier borrowers, who were previously denied credit by private banks, have a higher chance of obtaining credit from less cautious public banks, there may be evidence of differential selection at play for each type of bank.

In Brazil, a country where both private and state-owned banks command substantial portions of the loan market, researchers have explored the interactions between these two types of institutions. Moreover, state-owned banks have been used by governments to boost credit, through the belief that, as public banks slashed interest rates, other private banks would follow and it would increase competition in the market. Nonetheless, there is no evidence that state-owned banks compete aggressively with private ones. Sanches et al. (2018) shows that the probability that a private bank exists increases with the existence of a public bank, whereas the sign of this is inverse in the case of two private banks. The authors argue that there could be a complementarity between private and public financial institutions.

There is also evidence that the two types of banks operate differently. Coelho et al. (2013) explore the response of private banks to the entry of state-owned banks on local markets by using variation on market size and number and type of private incumbents. They show that public banks exert little competitive pressure on private banks, either because of product differentiation or cost differences. They argue that public banks may face higher costs, which could come because of higher delinquency rates or because public banks face a different clientele.

Ornelas et al. (2022) also show that banks behave differently with regards to the strategy employed with new and on-going borrowers. Private banks tend to attract new borrowers by offering lower interest rate and they increase the loan spread as the relationship evolves, a sign of the relevance of switching costs in the market. State-owned banks, on the other hand, decrease their spread as the relationship continues, once informational asymmetries are minimized.

The differences between private and state-owned banks imply that information asymmetries may vary between them. This variation could be due to differences in the types of clients they attract or disparities in their screening processes. In the Brazilian context, we focus on differences for one specific bank, Caixa Econômica Federal.

The main reason for that is that Caixa was primarly employed to increase the amount of personal loans as part of a governmental intervention in the credit market. The program called "Melhor Crédito" was carried out by the institution in the period from 2011 to 2014. In a paper using the microdata from Brazil in payroll loans, Garber et al. (2021) show that the increase in the loans from public banks from this period was focused mainly in less sophisticated public workers, as defined by an index that measures the degree of and income in the occupation. Although they show that there was a greater expansion to lower-income and high-risk borrowers, the share of default was not affected and their margin of adjustment was consumption reduction.

3
Data and statistics

3.1
Data

This paper combines data from different sources, all publicly disclosed by the Central Bank. These sources include data on news loans, average interest rates charged by financial institutions[1], number of branches and their geographical distribution, as well as accounting data from banks.

The first main database contains self-reported information on interest rates charged by financial institutions in Brazil, categorized by their tax ID numbers and across all loan categories. This data is reported daily, covering the mean of the past five days.

The second database is the System of Credit Registries (SCR Data), which provides quarterly information on the size of credit loans for independent institutions and financial conglomerates[2]. Credit loans are further categorized into different types, including modalities and the type of consumer (households or firms). These categories enable us to merge information of prices (interest rates) with quantities (loans). Additionally, this database includes information on the proportion of overdue loans in the credit portfolio, which serves as a proxy for default rates, and the number of credit operations for each credit line.

To combine the two datasets, we must make certain approximations since the information provided in both datasets differs. For instance, the data on interest rates reflects the average rate of new loans offered in each quarter[3]. On the other hand, data on quantities represents quarterly active credit portfolios for each institution, providing a measure of loan stock. To approximate the flow of loans, we observe the proportion of loans in each maturity and compute an approximate measure of the flow by excluding loans that expire within 90 days[4]. We then compute the difference between credit portfolios between two

[1]Available for independent financial institutions or the primary institution within conglomerates

[2]A financial conglomerate may include more than one institution, such as a bank or a cooperative union, so we consider only the bank with the largest credit portfolio

[3]The interest rate is reported on a daily basis, but we convert it into a quarterly measure by averaging its values over the respective period. We also observe three types of rates for payroll loans: public and private workers, as well as pensioners. In order to match to the loan data, we also take the average across these three categories.

[4]Loans that expire within 90 days in period $t-1$ serve as a proxy for period-specific amortizations. They are added to the difference between the credit portfolio at t and $t-1$

quarters. However, this measure does not take into account the proportion of loans that are paid in advance.

To compute the banks' market share, we follow the procedure implemented by Berry et al. (1995) and applied to the loan market in Brazil by Nakane et al. (2006). The consumer chooses to purchase one unit of "average banking loan", calculated as the sum of all new loans divided by the number of credit operations in each quarter. The market share is calculated as the sum loan units sold by a bank divided by the size of the potential market.

The potential market is given by the sum of sold goods and the outside good, defined as a measure of all loans which were not taken out in each period. We define the size of the outside good as the number of individuals that have a bank account, but did not take a loan in the period[5] multiplied by the average value of a loan[6].

Further information is sourced from the accounting data of financial institutions (IF Data) and branches (Estban). Both datasets include balance sheet information for institutions. IF Data provides more detailed quarterly accounting data for each institution, whereas Estban offers less detailed monthly data for all branches within an institution. Although the datasets encompass information about institutions' total credit portfolios, they lack differentiation between credit types. Hence, from these sources, we utilize variables related to deposits, assets, and branch counts, excluding credit-related ones.

We specifically filter data related to interest rates and the loan portfolio for payroll loans, which are one of the most significant categories of personal loans. Payroll credit involves automatic payment mechanisms tied to salary or retirement income, exclusively offered to individuals with stable earnings, such as employees with formal contracts, public servants, or pensioners.

While the data covers various types of financial institutions, our focus is primarily on commercial banks and multiple banks engaged in commercial activities. Although cooperatives[7] are prevalent in the market in number of branches, their aggregate credit portfolio represents a small share of total loans (Figure 9.4). Furthermore, as cooperatives operate differently from banks, with distinct revenue sources, higher interest rates, and a more localized presence, we exclude them from our analysis. Additionally, we exclude banks with infrequent credit loans to focus on institutions that actively engage in

[5]We make this approximation by subtracting the total number of loan operations in the quarter from the number of unique identification numbers (CNPJs) that have a bank account. This approach assumes that each person takes, at most, one loan per quarter.

[6]The total loan portfolio divided by the number of operations

[7]Credit cooperatives are defined as "Cooperativas de Crédito" and "Sociedade de Crédito, Financiamento e Investimento"

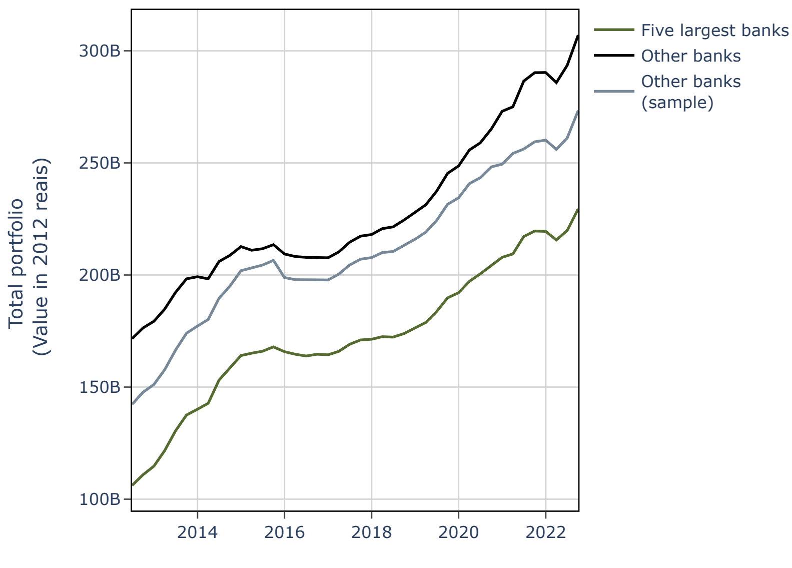
the market. This ensures that our analysis is centered on institutions with a significant presence in the lending industry. Consequently, our final sample comprises 39 institutions and the selected sample spans the period from 2013 to 2021.

3.2
Descriptive statistics

The importance of loans for individuals has increased in the period studied, with their share to total loans rising from 40% in 2012 to 56% in 2022 (Figure 9.2). Payroll loans represent approximately 18% of total loans for individuals, the second biggest portfolio after credit for habitations.

The average number of institutions that operate in this market for our relevant sample is approximately 30 banks. However, the payroll loan market exhibits high concentration: Caixa Econômica, Banco do Brasil, Itaú, Santander and Bradesco account for 75% of the total portfolio of the market. The graph below shows evolution of the total portfolio for the five largest banks[8] and the total of the market. The gray line represents the sum of the portfolio for the relevant institutions for this study, as described in the previous subsection.

Figure 3.1: Total portfolio



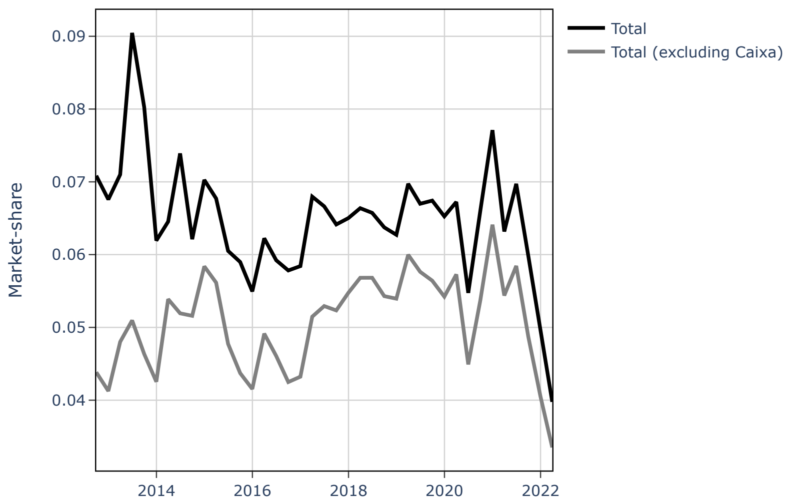
From the Figure 3.1, we can see that starting in 2020, the portfolio of loans outside our relevant sample has increased. This expansion is attributed to the rise in the number of cooperative banks and unions, coupled with the enlargement of existing ones, some of which have transitioned into commercial

[8]Portfolios per institutions are presented in Figure 9.3 in the Appendix

banks. However, despite this growth, it remains relatively small in comparison to the entire market.

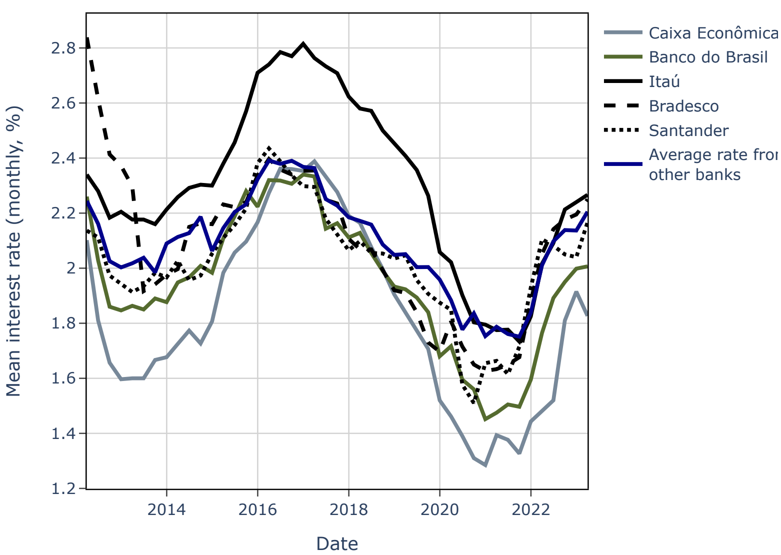
The two main variables of interest are the banks' market shares and interest rates. To compute the market shares, we derive a measure of new loans based on the portfolio variables, as described in the previous subsection. The Figure 3.2 presents the ratio of the sum of all new loans to the sum of new loans and the outside good.

Figure 3.2: Market share of the inside goods



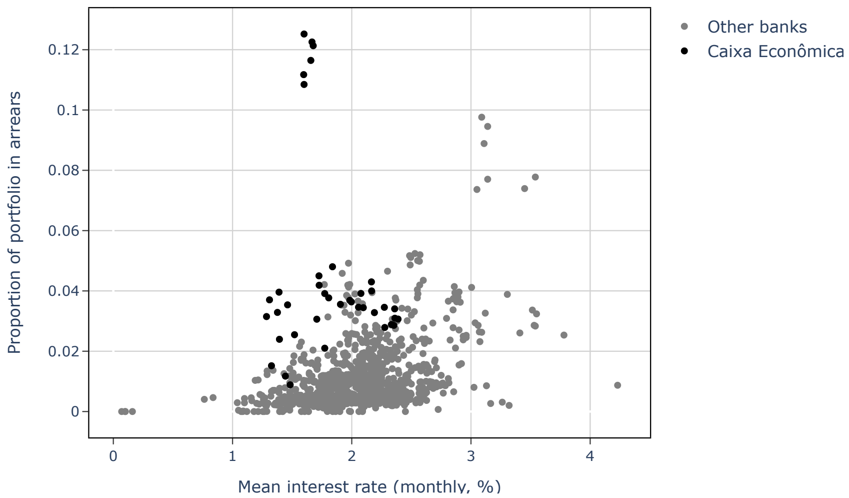
In Figure 3.3, the interest rates from the five largest banks are depicted, along with the average rate from the remaining banks in the sample. There is a low dispersion of rates for this type of credit and they closely track the interbank interest rate, Selic.

Figure 3.3: Interest rates



The delinquency rates, which represent the level of risk for banks for the relevant sample, lie in the range of 0 to 12%. The correlation between interest and default rates is low, but positive, which might indicate that banks that charge higher interest rates are the ones with higher costs from default, though the sign of the causal effect cannot be identified.

Figure 3.4: Interest rates vs. Delinquency rates



From the figures above, we can see how Caixa Econômica differs from other banks in its pricing and lending behavior. In Figure 3.2, the credit expansion between 2013 and 2014 can be attributed primarily to Caixa, as part of a government driven policy, whereas the other banks did not exhibit a similar pattern. During the same period, Caixa exhibited the highest default rate in the market, as evidenced by the scatter points in Figure 3.4, where its delinquency rates reached approximately 12%. Furthermore, Caixa consistently maintains the lowest interest rate among the largest banks (as shown in Figure 3.3) throughout the entire period, while holding the second-highest loan portfolio (Figure 9.3).

Table 3.1 presents some summary statistics on the mean values for interest rates, average loans and other bank characteristics.

Table 3.1: Summary statistics

	Mean	Std	Min	25%	50%	75%	Max
Interest rates (%)	2.07	0.41	0.76	1.80	2.07	2.31	4.23
Market-shares (%)	0.24	0.44	0.00	0.01	0.04	0.16	3.95
Average loan (R$, M)	883	1643	0.03	22	148	622	13602
Number of banks	27.60	2.09	20.00	27.00	27.00	29.00	32.00
Branches (log)	3.74	2.91	0.00	0.69	4.14	5.69	8.62
Age	50.70	24.26	4.00	31.00	50.00	71.00	115.00
Default rate (%)	1.23	1.41	0.00	0.43	0.80	1.49	12.52
Basic interest rate (%)	0.19	0.05	0.06	0.17	0.19	0.23	0.25

Rates are expressed in monthly frequency

Average loans are expressed in million reais, in 2012 values

3.3 Market shares and delinquency rates

In this section, we aim to examine whether fluctuations in banks' market shares are correlated with changes in default rates. Such a correlation would suggest that as banks expand their credit portfolios, new borrowers are being included in the pool, potentially impacting default rates.

We estimate the following panel-date specification, where j is the subscript for the financial institution and t represents the quarter.

$$default_{jt} = \alpha_0 + \alpha_1 share_{jt} + \alpha_2 X_{jt} + \xi_j + \nu_t + \epsilon \tag{3-1}$$

where $default_{jt}$ is the ratio of the loan portfolio in arrears for more than 90 days to the total portfolio, $share_{jt}$ is the proportion of new loans for each bank in the quarter (market share) and X_{jt} is the vector of banks' characteristics, such as number of branches and age of the institution. We add bank and time fixed effects, denoted as ξ_j and ν_t.

Table 3.2: Default rates and market shares

	Dependent variable: Default rate		
	(1)	(2)	(3)
Intercept	0.0228	0.0273	-0.0084
	(0.5866)	(0.6938)	(-0.1988)
Market-share	1.5770***		
	(2.7754)		
Market-share (1 lag)		1.7439***	
		(2.9930)	
Market-share (2 lags)			1.4553**
			(2.5688)
Branches	-0.0001	1.281e-05	8.996e-05
	(-0.2453)	(0.0217)	(0.1583)
Age	-0.0003	-0.0004	0.0003
	(-0.3615)	(-0.4957)	(0.3986)
Effects	Entity	Entity	Entity
	Time	Time	Time
No. Observations	998	988	976
R-squared	0.1252	0.1666	0.1281
F-statistic	43.883	60.653	43.993

T-stats reported in parentheses, robust standard errors

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

There is a significant correlation between the variables, which indicates that there is sufficient variation in the data for the identification of the model. The market share is positively correlated with the delinquency rates, even though the total credit portfolio expands when market share increases.

Additionally, the strongest correlation between the delinquency rate and the market share is with its first lag, shown in column (2). This suggests that the effect of the expansion of credit on the default rate typically becomes apparent after one quarter. While this is not an exact measure, given the limitations in tracking loans to their default periods, we will rely on this approximation.

4
Model

First, we present the demand estimation framework used to determine the parameters that govern the utility of agents in each market. Assume we observe T markets over the period t=1,...,T, where individuals i=1,...,I decide which banks to purchase loans from. The utility of a consumer i who chooses to take a loan from financial institution j in period t is determined by the loan's average interest rate p_j and the bank's characteristics x_j, as well as the consumer's type. Each individual is associated to a type θ which represents their probability of loan default, so high-risk borrowers present θ closer to 1. The distribution of types in the population is given by the probability density function $f(\theta)$, with mean μ and standard deviation σ.

$$u_{ijt} = \underbrace{\alpha \cdot p_{jt} + \beta \cdot x_{jt} + \xi_j}_{\delta_{jt}} + \nu_t + \gamma\theta_i + \epsilon_{ijt} \tag{4-1}$$

The firm and year fixed-effects, represented by ξ_j and ν_t, capture preferences for specific banks and changes in the macroeconomic cycles. The mean utilities for borrowing from each bank are given by δ_{jt}. The mean-zero stochastic term ϵ_{ijt} represents individual's preferences and it follows an Extreme Type I distribution.

Besides taking a loan, the consumer also has the option of not purchasing the good from the selected institutions, which is referred to as the "outside good". The outside option is defined as the choice of not taking a loan or obtaining it from a financial institution other than a commercial bank, such as credit unions or cooperative banks. Its utility is given by:

$$u_{i0t} = \epsilon_{ijt} \tag{4-2}$$

The risk type θ of borrowers affects only the utility of the inside goods, in other words, it influences the decision to borrow or not. We account for the information asymmetry in the demand estimation framework through the parameter γ, which measures the impact of θ on the utility. A positive γ points to the presence of adverse selection, as riskier consumers obtain more utility when taking out loans. On the other hand, advantageous selection might exist in this market if γ is negative, implying that those more likely to seek credit present lower risk to banks.

To provide a tractable test for such asymmetry, we assume that the informational component affects all institutions equally, so γ is the same for all banks. The goal is to identify a mean measure for the asymmetry across the market, since the idiosyncratic differences in risk-taking behavior for financial institutions are not represented in this model.

We also aim to represent consumer heterogeneity in unobservable characteristics through the inclusion of the θ types in the model. This idea is closely related to random coefficients introduced by Berry et al. (1995) (BLP). In BLP's model, the price is the same for all individuals, but they react differently to it, so the coefficient has a distribution which is estimated. In our model, consumers differ in their level of risk θ, which also follows a distribution, whereas the coefficient γ is common to borrowers and banks. Therefore, we allow some dimension of the choice motivation to vary across individuals in order to relate it to a risk-related measure for banks. This approach allows us to extract information on unobservable characteristics only by exploring the variation at the aggregate level, but in an alternative way to BLP.

In the model, the consumer chooses to borrow from the institution that provides them with the highest utility. Following the multinomial logit model (McFadden, 1973), the probability a consumer chooses an option among all available can be analytically calculated because of the hypothesis made upon the distribution of ϵ [1].

We introduce a variation in the logit model: instead of calculating the total market share of a institution, we compute the market share conditional on a risk type θ:

$$s_{jt}(\theta) = \frac{e^{\delta_{jt}}}{e^{-\gamma\theta-\nu_t} + \sum e^{\delta_{nt}}} \tag{4-3}$$

According to the Law of Iterated Expectations, the unconditional market share for each bank is calculated as a function of the fraction of the population which is of type θ and their corresponding market share for each bank:

$$s_{jt} = \mathbb{E}_\theta[s_{jt}(\theta)] = \int_0^1 s_{jt}(\theta) f(\theta)\, d\theta \tag{4-4}$$

The total demand for credit Q_t is then given by the sum of the market shares of all n institutions operating in the market.

$$Q_t = \sum_{j=1}^{n} s_{jt} \tag{4-5}$$

The ex-ante type of the consumers influences not only their decision to take on a loan, but also their success in repaying the loans. Since each borrower has a default probability denoted as θ, the overall loan default rate for a bank is determined by the average θ among its consumers. The delinquency rate for each bank, denoted as a, is determined by calculating the expected value of

[1]The demonstration is shown in the Appendix

the parameter θ within its portfolio, so it is weighted by the relative market share of the bank conditional on the risk type.

$$a_{jt} = \mathbb{E}_{\theta}[s_{jt}(\theta) \cdot \theta] = \int_0^1 \frac{\theta \cdot s_{jt}(\theta) \cdot f(\theta)}{s_{jt}} \, d\theta \tag{4-6}$$

Using Equations 4-5 and 4-6, we create a system of equations that relates levels of banks' market shares to their default rates. This relationship is contingent upon an assumed distribution of risk types within the population, which affects the borrowing decisions and their outcomes. Once the demand parameters' levels and the size of the outside good are defined, the system is estimated in order to recover the parameters γ and the mean and standard deviation of the distribution of θ.

When a bank increases its lending activities, it expands its pool of borrowers, which includes marginal consumers who were not part of the previous period's borrower base. This expansion has implications for the delinquency rates of that bank in the subsequent period, as the newly included consumers reveal their ex-ante risk types through ex-post measures. As argued by Meza and Webb (1987), in the case of advantageous selection, the marginal borrower tends to represent the worst risk type among the pool of borrowers, whereas in situations of adverse selection, the marginal borrower tends to be the safest type among the borrowers.

The relationship between the demand and default rate of marginal consumers is used to identify the selection parameter of interest in the paper. When selection is adverse, the market share among high risk borrowers is significantly higher and credit expands relatively more among low-risk borrowers. On the other hand, when selection is advantageous, credit expansions lead to the inclusion of riskier marginal individuals to the pool of borrowers. The bank's default rate, a weighted average of different risk types borrowing from each bank, indicates the prevailing selection type. In scenarios with advantageous selection, the projected default rate increases with credit expansion, while in adverse selection situations, it decreases.

4.1 Alternative model

State-owned banks exhibit distinct characteristics compared to private banks and play a significant role in capturing a substantial portion of the market share for loans. As an alternative specification, we modify the model to incorporate a distinct effect of the ex-ante risk component for one of the largest banks in Brazil, namely Caixa Econômica Federal.

We hypothesize that the utility experienced by consumer i, who chooses to borrow from Caixa, differs from the utility described in Equation 4-1, primarily due to a variation in the γ parameter. Hence, we denote this modified parameter as γ_c.

$$u_{ict} = \underbrace{\alpha \cdot p_{ct} + \beta \cdot x_{ct} + \xi_c}_{\delta_{ct}} + \nu_t + \gamma_c \theta_i + \epsilon_{ijt} \tag{4-7}$$

Following the same steps from the previous subsection, we compute the market share conditional on risk type θ for for Caixa and for the rest of the banks, as it is done in Equation 4-3.

The market share conditional on the risk type θ for all banks j except Caixa and for Caixa are given by:

$$s_{jt}(\theta) = \frac{e^{\delta_{jt}}}{e^{\delta_c + (\gamma_c - \gamma)\cdot\theta} + e^{-\gamma\theta - \nu_t} + \sum e^{\delta_{nt}}} \tag{4-8}$$

$$s_{ct}(\theta) = \frac{e^{\delta_c + (\gamma_c - \gamma)\cdot\theta}}{e^{\delta_c + (\gamma_c - \gamma)\cdot\theta} + e^{-\gamma\theta - \nu_t} + \sum e^{\delta_{nt}}} \tag{4-9}$$

The total quantity of credit extended by each bank and Caixa, denoted as s_{jt} and s_{ct}, is the expected value of the conditional market share with respect to risk type.

$$s_{jt} = \mathbb{E}_\theta[s_{jt}(\theta)] = \int_0^1 s_{jt}(\theta) f(\theta)\, d\theta \tag{4-10}$$

$$s_{ct} = \mathbb{E}_\theta[s_{ct}(\theta)] = \int_0^1 s_{ct}(\theta) f(\theta)\, d\theta \tag{4-11}$$

The total demand for credit Q_t is given by the sum of the market shares of $n-1$ institutions operating in the market, with Caixa being excluded.

$$Q_t = \sum_{j=1}^{n-1} s_{jt} \tag{4-12}$$

The system of equations is composed by the total demand for credit excluding Caixa (Equation 4-12), Caixa's market share (Equation 4-11) and banks' default rates (Equation 4-6). We recover the parameters δ_c and γ_c, in addition to the existing parameters γ, μ, σ, and the time fixed effects ν_t.

5
Econometric specification

The estimation is conducted in two steps. In the first step, our goal is to estimate the bank-market specific constants, denoted as δ_{jt} in Equation 4-1. The second stage uses the predicted market shares from the first stage to estimate a system of two equations: demand for credit lines and default.

First Step Estimation: In order to estimate the mean level of utility across banks, we define the main characteristics that influence the consumers' decision. The estimation then faces two main challenges. First, according to Equation 4-1, the consumer's utility depends not only on banks' characteristics, but also varies with the consumer type, which is unobservable for the econometrician. To address this issue, we rely on a simplifying hypothesis that enables us to carry out the estimation using the available data. Secondly, the interest rate is potentially correlated with other factors that influence the decision to borrow, so the endogeneity issue must be addressed with the use of instruments.

Bank characteristics: The borrower's utility is influenced by two variables: the number of branches and the age of the financial institution. Consumers often access their bank accounts through physical branches, especially in municipalities where only a few banks operate. Thus, the number of branches can be a relevant indicator of accessibility and convenience for customers.

The age of the institution might have an ambiguous effect. Older banks tend to be more established and well-known, which may lead to a perception of greater reliability among customers. However, young people show a preference for digital banks and may perceive them as more innovative and technologically advanced.

Furthermore, consumers may have specific reasons for preferring a particular bank, such as its geographical location or factors beyond their control. For example, in Brazil, companies often establish agreements with banks to direct their employees' paychecks through them. This arrangement can lead to increased usage of the bank's services by the consumers, but it may not necessarily indicate superior service quality or individual preferences for that bank. Nevertheless, once a consumer starts using a specific bank, switching to another becomes more challenging due to significant switching costs (Silva and Lucinda (2017), Ornelas et al. (2022)). As a result, individuals tend to maintain their preference for the same bank over extended periods. To account for the effects of time-invariant bank-specific preferences, we include bank fixed effects. In addition, time fixed effects are included to help control for macroeconomic fluctuations that affect all institutions in a given quarter.

Once the bank's characteristics are defined, we can proceed to estimate the demand parameters.

Identification: To identify the parameters α, β and ξ_j and compute the mean utilities for all financial institutions that provide loans, we must address the unobservable variable θ in the borrower's utility. To achieve this, we conduct a straightforward algebraic manipulation of the market share equations that effectively removes the influence of this variable.

We conduct a normalization on one of the J inside goods, bank k[1]. We compute the difference of the market share, given by Equations 4-3 and 4-4, between banks j and k, which yields the following expression:

$$\log(s_{jt}/s_{kt}) = \delta_{jt} - \delta_{kt} \tag{5-1}$$

Since the parameter γ equally affects the utility of consumers that purchase loans, it will not affect the decision of consumers regarding the option chosen. We can rewrite the equation as a function of the variables that affect the consumers' utility:

$$log(s_{jt}) - log(s_{kt}) = \alpha \cdot (p_{jt} - p_{kt}) + \beta \cdot (x_{jt} - x_{kt}) + \xi_{jt} - \xi_{kt} \tag{5-2}$$

The second equation involves observable variables in both sides, so δ is estimated using the available data, once we set ξ_k to zero.

However, the price (or price difference) as an independent variable can be potentially correlated with unobservable factors that influence the loan quantity. Unobservable characteristics to the econometrician can influence the borrower's preference for a specific bank and might be correlated to the bank's interest rate. In the context of personal credit, individuals may prefer a particular bank branch based on its services, leading them to choose that bank as their primary institution. Consequently, they may receive better fees and rates.

To address this endogeneity issue, instrumental variables are necessary. We use instrumental variables associated to the banks' lending costs, as it commonly done in the literature, such as the basic interest rate, the proportion of funding resources and interest rates of other types of loans. The aim is to provide exogenous shocks to the price difference between each bank and the

[1]In the standard logit model, the normalization is usually performed with the outside good in a market with J+1 options

reference bank and to induce variation in the supply of loans. Next, we justify the chosen instruments.

The primary cost for banks in our analysis is the basic interest rate, known as Selic. This rate represents the weighted average interest rate of overnight interbank operations and serves as the main accounting and opportunity cost for banks when granting loans. Over the period from 2013 to 2021, we observe significant fluctuations in the Selic rate, reflecting Brazil's monetary cycles. Although it is one of the main macroeconomic variables, we assume that the effect it has on the market share happens through the changes in the bank's interest rate. Since the credit market operates at the local level, we assume branches are price-takers and they do not contract or expand the level of loans in the short-term. While Selic significantly influences a bank's interest rate, its impact on the interest rate difference between each bank and the reference bank, our dependent variable, relies upon the heterogeneity of individual banks' reactions to variations in this rate.

We also use the proportion of total deposits to financial assets as an instrument, a measure which is similarly used by Crawford et al. (2018). Deposits are one of the main sources of funds used to give out loans in banks, so an increase in the proportion of deposits is associated to a higher availability of funds and lower lending costs. We assume that deposits are mainly driven by firms and households, a different pool of clients than the ones that demand personal credit loans. We also interact deposits with basic interest rate (Selic), since it is expected that there is a differential effect when there is a high cost of credit. When Selic is high, institutions that have a big share of deposits have a higher opportunity cost to offer loans, so we expect interest rates to be positively correlated with this term. Another significant source of variation in our analysis is the working capital interest rates. These rates represent the opportunity cost of loans or may be influenced by higher costs incurred by the bank. It is expected that higher interest rates for firms would be positively correlated with rates for personal credit. We justify the exclusion restriction in this context by considering that the demand for working capital loans is primarily driven by firms and is generally of a short-term nature. As a result, it is not correlated with the demand for personal credit, which is associated with the needs of households.

As a second set of instruments, we use the lagged variable of the difference in interest rates between two banks. The rationale behind this instrument relies on the persistence observed in interest rates, which reflects a bank's lending capacity in the subsequent period.

The regression for the relevance of the instruments is reported in Table

9.1. We will use the specifications (4) and (5) in the following tests.

After estimating the demand parameters that govern the utilities associated with the inside goods, we proceed with the second step of the estimation.

Second Step Estimation: In this step, we consider a market with $J + 1$ goods, which includes the outside good. Our model considers that banks provide loans to consumers with varying levels of risk and observe delinquency rates from these consumers, an ex-post measure of their risk profile. Consequently, in our analysis, it is essential to have sufficient variation in loan shares and a significant correlation between loans and the proportion of overdue loans for each period, and it is confirmed by Table 3.2.

We define two equations representing the total demand for loans and the default rate of each bank, utilizing the demand parameters obtained in the first step. The Equations 4-5 and 4-6 are then estimated in a system using the Nelder-Mead simplex algorithm. We match the observed market share and default patterns to those generated by our model, which are derived based on a constant risk distribution from the population. We then recover the parameters of interest γ and the mean and standard deviation of the risk type distribution, as well as time fixed effects.

In the model, the total demand in the market is a function of the mean utilities of each bank-market, the effect of the selection parameter, the distribution of risk types and the size of the outside good.

The delinquency rate is given by the expected value of the risk types for a bank. From the data, also based on the results shown in Table 3.2, we choose to use one period forward delinquency rate as a measure of default in the model. This happens because the ex-post measure of delinquency can only be observed after the loan has been conceded. Since we do not have information about the time correspondence of loans and its defaults, we consider an average value of one quarter.

Moving forward, we address certain challenges in the estimation process, such as the outside good definition and the choice of the shape of the risk distribution, and elaborate on the decisions made to address these challenges.

In the model, the total demand for loans relies on the size of the outside good. However, the outside good is a value we do not observe, since it refers to all operations that did not occur, so we turn to approximations to determine the size of this market. This measure plays a crucial role in identifying the γ parameter. To address this issue and ensure the robustness of our findings, we conduct an analysis of this aspect of the model in the Appendix. We explore the impact of different sizes of market shares on our estimated model. By varying the share of the outside good, we can assess the sensitivity of our results and determine the extent to which they are influenced by the measure of the outside good. We acknowledge that we cannot accurately identify the shape of this distribution for the whole population from the available data. However, the relevant effect studied through the model is primarily driven by the behavior of marginal consumers.

The distribution of risk types follows a Logit-normal distribution[2]. The choice for such distribution stems from the fact that it is flexible enough to allow for the case of a bimodal distribution centered around 0 and 1, as well as a normal distribution. This versatility is desirable because contingent on the size of the outside option, the model accommodates a greater prevalence of type 1 consumers.

Robustness test: As an additional test to enhance the robustness of our analysis, we perform a second model estimation using data from only one bank. This approach allows us to assess whether there the sign of selection is the same across all banks in this specification and if there is a disparity in the magnitude of the parameter γ between Caixa and other banks.

The first step is conducted in the same way: the multinomial logit is estimated for J banks, excluding the outside option. The predicted market shares are calculated based on the demand parameters estimated.

In the second step, the observed empirical patterns for the demand for a given bank and their default rate are matched to those generated by the model. Equations 4-4 and 4-6 are estimated in a system using the Nelder-Mead simplex algorithm in order to recover the parameter γ and the quarter-fixed effects. In this specification, we fix the moments of the distribution of θ (μ and σ) as the values presented in the first estimation in order to reduce the number of parameters estimated.

[2]The probability distribution of the logistic transformation of a normally distributed variable

6
Results

6.1
First step

In the first step, we estimate the regression of the banks' market shares on their prices and characteristics, represented by Equation 5-2, using Ordinary Least Squares and Instrumental Variables. All variables are measured as deviations from those of bank k and the sample of banks includes $J-1$ banks.

For column (2), we use the set of instrumental variables related to the banks' costs: Selic (basic interest rate), interest rates for firms, deposits and its interaction with Selic. For column (3), we use only the lagged price of each bank as a source of exogenous variation.

Table 6.1: Multinomial logit model estimation

	OLS	IV	
		Banks' costs	Lagged prices
	(1)	(2)	(3)
Interest rates	-0.46***	-1.15***	-0.77***
	(0.09)	(0.38)	(0.14)
Branches	0.10	0.19	0.12
	(0.06)	(0.08)	(0.08)
Age	-0.01	-0.01	-0.02
	(0.05)	(0.05)	(0.04)
Intercept	-0.24	-0.87	-0.86
	(2.64)	(2.68)	(2.59)
Effects	Bank	Bank	Bank
N	962	962	952
F-stat	454.47	42303.82	44465.05
Adj. R^2	0.95	0.95	0.95

Standard errors reported in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Price coefficients are negative and significant at the 1% level for both sets of instruments, and they exhibit larger magnitudes compared to the OLS estimates. Bank characteristics, such as branches and age, are not relevant to explain the market shares. However, bank fixed effects capture the preference for such features.

The coefficients for the first step of the alternative model are presented in Table 9.3 in the Appendix.

6.2 Second step

In the second step, we estimate the system of equations as outlined in Equations 4-5 and 4-6. The resulting estimates for the parameter γ are -43 and -32, corresponding to the two sets of instruments used. Our primary focus is on the sign of this parameter rather than its specific numerical value. Hence, no interpretation of the magnitude is conducted. The parameter points to the presence of advantageous selection, since higher-risk consumers derive a lower utility from taking out a loan.

The model correctly captures most of the variation of the market share of all inside options and the mean value of the delinquency rate. One drawback from the theoretical structure of the model is that it implies that default rates are the same for all institutions in the period, so it can never fully capture the idiosyncrasies of the banks.

Figure 6.1: Total demand for loans

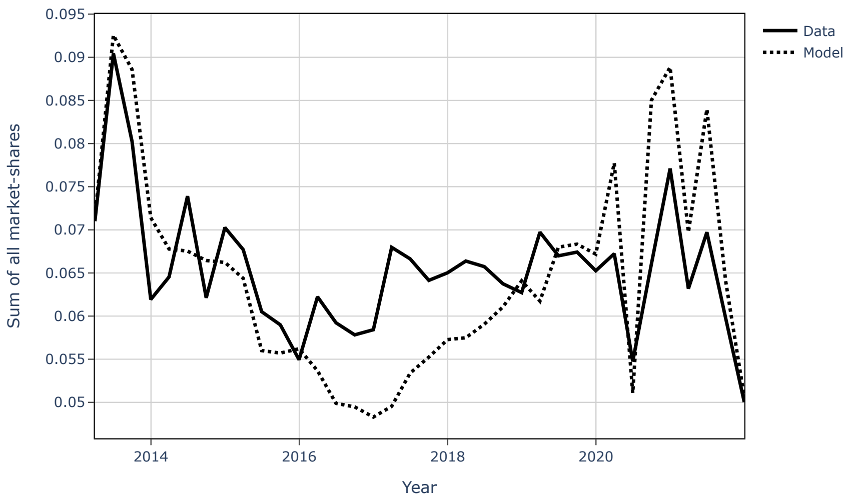
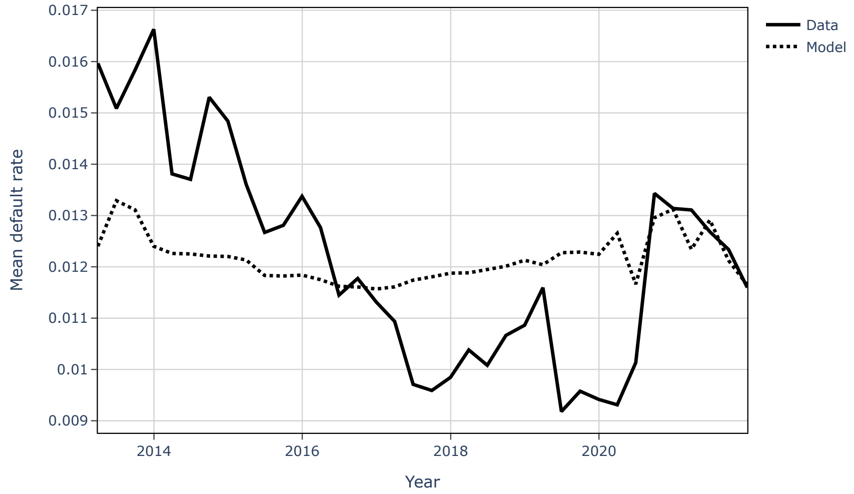
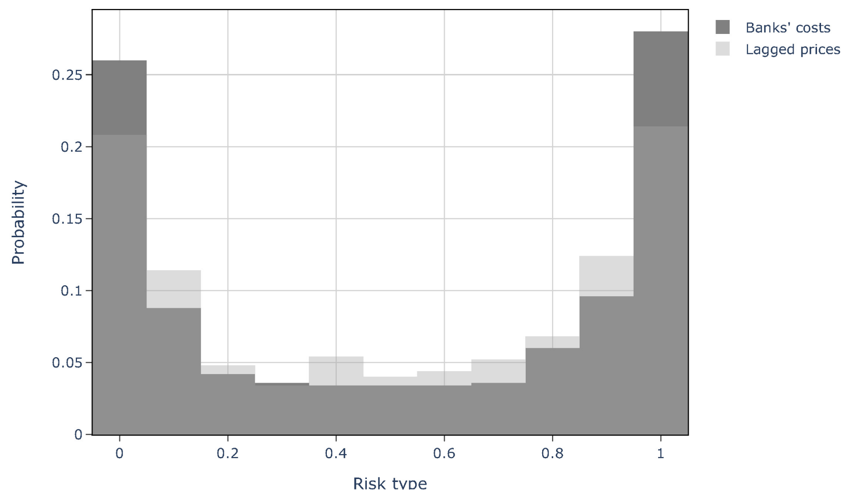


Figure 6.2: Average default rate



In Figure 6.3, we illustrate the logit-normal shaped distribution estimated for the risk type θ in the entire sample using both sets of instruments.

Figure 6.3: Estimated distribution of risk types



The proportion of individuals with θ equal to one is generated by the model to account for the substantial percentage of consumers, nearly 92%, who opt for the outside good in the market. However, a share of these individuals who are perceived as high-risk consumers may actually have no inclination to borrow in this market and repay their loans, even if, in the event they demanded a loan and applied for it, they could potentially be low-risk borrowers. We acknowledge that this distribution does not represent the risk profile of the entire population but rather serves as a means to provide context for assessing the decisions of marginal consumers who contemplate borrowing in this market.

Although the model provides a good fit, the total demanded loans hides opposing forces: the high level of loans from 2013 is mainly driven by Caixa, as well as the increase in loans from end of 2014 to 2016, whereas the rest of banks experienced the contrary movement. By separately modeling Caixa, we are able to better fit the model. As a result, the parameter of interest is higher than before, -7.4 or -17.4, while γ_c has increased even further to -4.2 or -5.7. The estimated parameters are presented at Table 6.2[1].

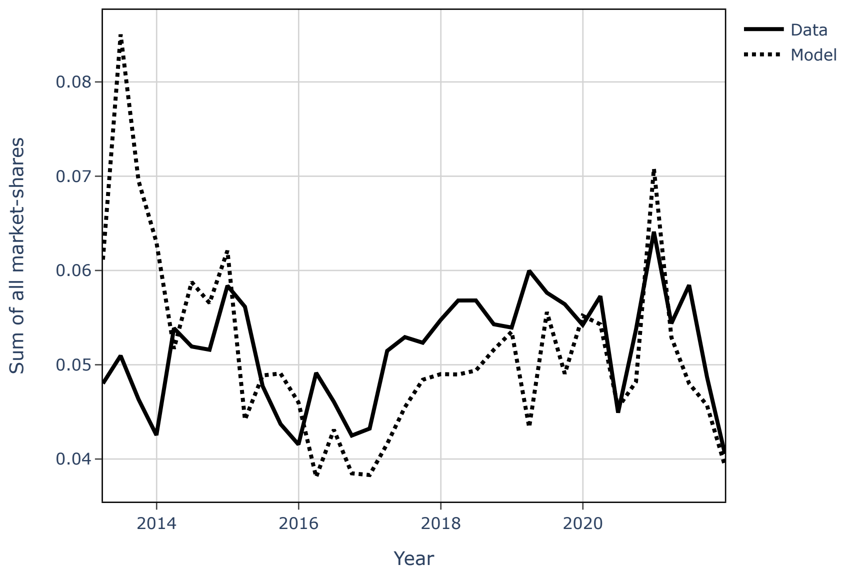
Table 6.2: Parameters estimates

	Homogeneous effect		Diff. Caixa effect	
	Banks' costs	Lagged prices	Banks' costs	Lagged prices
γ	-43.11	-31.93	-7.35	-17.39
γ_c			-4.19	-5.67
μ	0.52	0.52	0.53	0.53
σ	0.39	0.47	0.47	0.44

All coefficients are significant at 1%

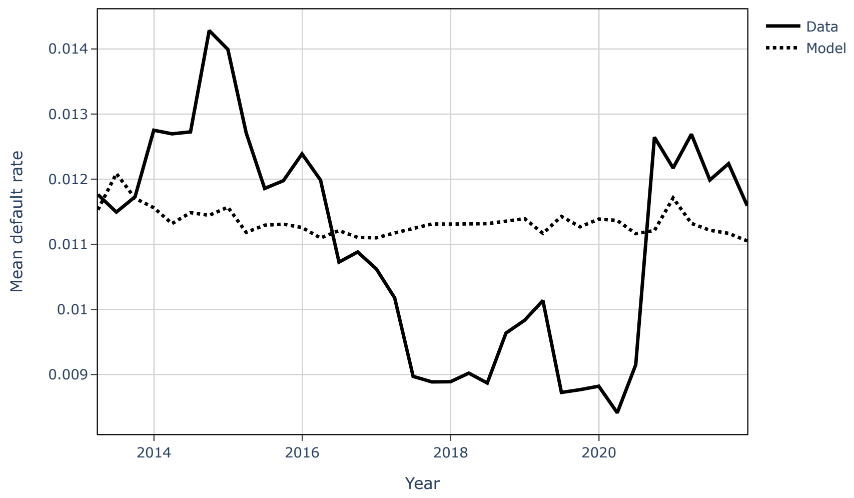
The model fit improves once Caixa is excluded from the sample, as shown in Figures 6.4 and 6.5.

Figure 6.4: Total demand for loans (excluding Caixa)



[1]Following Crawford et al. (2018), standard errors are computed with 500 bootstrap replications.

Figure 6.5: Average default rate (excluding Caixa)



Figures 6.6 and 6.7 show the demanded quantity and default rate of Caixa. The main driver for the increase in loans for Caixa in the beginning of the period was the government intervention (Garber et al., 2021), which began in 2012. It was accompanied by a higher default rate, even before the 2014 recession started.

Figure 6.6: Demand for loans in Caixa

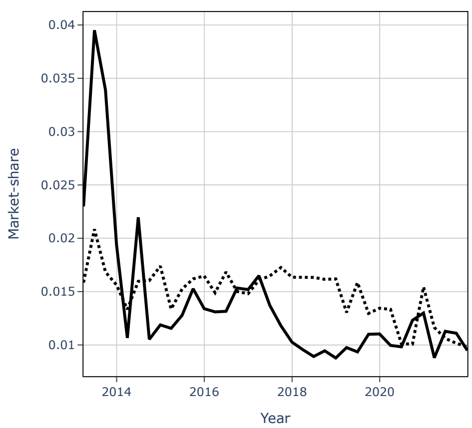
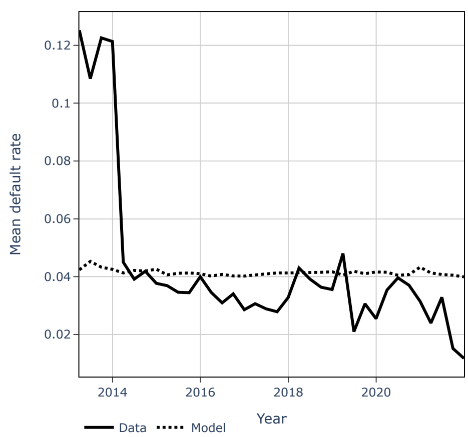


Figure 6.7: Default rate in Caixa



The model captures the relationship between the market share increase and the higher proportion of high-risk borrowers in Caixa. This is in agreement with (Garber et al., 2021), which shows that higher risk individuals were the targeted consumers from the credit expansion led by state-owned banks in Brazil. Also, the sustained higher level of default even after the government intervention is discontinued influences the estimation of the parameter γ_c.

6.3 Robustness test

As a robustness test, we fix the distribution of risk types and estimate the model separately for a subset of banks that operate in every quarter of the sample. The results are presented in the Table 6.3.

Table 6.3: Robustness test: Estimation of parameter γ for each bank

Bank	γ **estimated**
Caixa	-12.1
Santander	-22.5
Itaú	-60.7
Bradesco	-84.4
Industrial do Brasil	-85.4
Banco do Brasil	-96.5
Banco de Brasília	-150.4
Banestes	-156.8
Bancoob	-184.9

The parameter γ for Caixa presents the lowest value in magnitude, consistent with our prior findings.

It is worth to note that Banco do Brasil, the state-owned bank with the largest credit portfolio, presents a value of γ higher in magnitude than large private banks, such as Itaú and Santander. This suggests that the result from Caixa might be driven mainly by the period of the governmental intervention, since Banco do Brasil did not participate in the credit programs focused on payroll loans.

7
Conclusion

This paper studies the payroll loan market in Brazil. We aim to understand whether this market presents information asymmetries and whether they lean towards adverse or advantageous selection. We combine multiple sources of publicly firm-level data and we develop a model which allows us to extract information on consumers' unobservable risk-related characteristics. We extend the demand estimation framework proposed by McFadden (1973) to include the effect of consumer risk heterogeneity on lending decisions.

Our results are consistent with the existence of advantageous selection in the payroll loan market, indicating that more secure consumers tend to opt for loans. We find a selection parameter γ that reduces the utility of borrowing as the probability of default of the consumers increases. Although we model the effect through the demand side, the effect may reflect not only consumer preferences, but also the banks' ability to choose low-risk borrowers, since the purchase of a loan involves the agreement of the bank.

Many factors support these results for the credit for payroll loans in Brazil. Bank lending costs, which include consumer attraction, screening, loan design, and monitoring, require significant technology and labor investments, making them burdensome for banks. Along the operating costs, there is a latent risk of default. Moreover, in the Brazilian context, institutional and legal challenges further hinder the recovery of guarantees. As a result, banks exercise caution in selecting borrowers. Consumers of payroll loans, drawn from a sample of individuals with a relatively stable income, cannot easily default their loans, so they may exhibit a higher degree of risk aversion and greater diligence when making borrowing decisions.

Our results, however, do not account for market conduct effects. For South Korea, Kim (2017) shows that an uncompetitive market is needed to support advantageous selection in the credit market, which is further to be investigated in the Brazilian market. Moreover, we cannot separate the effect of consumer preferences from bank monitoring, which is important to determine how policy could improve the credit conditions.

The research also aims to adapt the proposed model to include a selection parameter specific to Caixa Econômica, since the state-owned bank behaved differently from other banks in this market during the studied period. As a result, we find a smaller selection parameter in magnitude, indicating that riskier borrowers find it more advantageous to get loans from Caixa compared to other banks. This might imply that Caixa is less concerned about borrowers' risk levels and the potential extra expenses for their operations than other banks in the market.

8
Bibliography

Adams, W., Einav, L., and Levin, J. (2009). Liquidity constraints and imperfect information in subprime lending. *American Economic Review*, 99(1):49–84.

Agarwal, S., Ambrose, B. W., Chomsisengphet, S., and Liu, C. (2006). An empirical analysis of home equity loan and line performance. *Journal of Financial Intermediation*, 15(4):444–469.

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Ausubel, L. M. (1999). Adverse selection in the credit card market.

Banerjee, A. V. and Duflo, E. (2014). Do Firms Want to Borrow More? Testing Credit Constraints Using a Directed Lending Program. *The Review of Economic Studies*, 81(2):572–607.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):841–890.

Cardon, J. H. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. *The RAND Journal of Economics*, 32(3):408–427.

Chiappori, P. and Salanie, B. (2000). Testing for asymmetric information in insurance markets. *Journal of Political Economy*, 108(1):56–78.

Coelho, C. A., De Mello, J. M., and Rezende, L. (2013). Do public banks compete with private banks? evidence from concentrated local markets in brazil. *Journal of Money, Credit and Banking*, 45(8):1581–1615.

Cohen, A. (2005). Asymmetric information and learning: Evidence from the automobile insurance market. *The Review of Economics and Statistics*, 87(2):197–207.

Crawford, G. S., Pavanini, N., and Schivardi, F. (2018). Asymmetric information and imperfect competition in lending markets. *American Economic Review*, 108(7):1659–1701.

Davidoff, T. and Welke, G. (2004). Selection and moral hazard in the reverse mortgage market. *SSRN Electronic Journal*.

Dell'ariccia, G., Friedman, E., and Marquez, R. (1999). Adverse selection as a barrier to entry in the banking industry. *RAND Journal of Economics*, 30(3):515–534.

Dick, A. A. (2008). Demand estimation and consumer welfare in the banking industry. *Journal of Banking Finance*, 32(8):1661–1676.

Edelberg, W. (2004). Testing for adverse selection and moral hazard in consumer loan markets. (2004-09).

Fang, H., Keane, M., and Silverman, D. (2008). Sources of advantageous selection: Evidence from the medigap insurance market. *Journal of Political Economy*, 116(2):303–350.

Finkelstein, A. and McGarry, K. (2006). Multiple dimensions of private information: Evidence from the long-term care insurance market. *American Economic Review*, 96(4):938–958.

Finkelstein, A. and Poterba, J. (2004). Adverse selection in insurance markets: Policyholder evidence from the u.k. annuity market. *Journal of Political Economy*, 112(1):183–208.

G1 Globo (2012a). Após bb e caixa, bancos privados avaliam redução de taxas de juros.

G1 Globo (2012b). Dilma critica altas taxas de juros e diz que bancos têm lógica perversa.

G1 Globo (2013). Caixa irá oferecer r$130 bi em crédito para pessoa física em 2013.

Garber, G., Mian, A. R., Ponticelli, J., and Sufi, A. (2021). Consumption smoothing or consumption binging? the effects of government-led consumer credit expansion in brazil. Working Paper 29386, National Bureau of Economic Research.

Gunnsteinsson, S. (2020). Experimental identification of asymmetric information: Evidence on crop insurance in the philippines. *Journal of Development Economics*, 144:102414.

He, D. (2009). The life insurance market: Asymmetric information revisited. *Journal of Public Economics*, 93(9):1090–1097.

Karlan, D. and Zinman, J. (2009). Observing unobservables: Identifying information asymmetries with a consumer credit field experiment. *Econometrica*, 77(6):1993–2008.

Kim, M. (2017). Multidimensional heterogeneity and the nature of advantageous selection in the consumer credit market.

Kim, M. and Yoon, J. (2022). Asymmetric information and imperfect competition: Evidence from the korean personal loan market.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pages 105–142.

Meza, D. D. and Webb, D. C. (1987). Too much investment: A problem of asymmetric information. *The Quarterly Journal of Economics*, 102(2):281–292.

Nakane, M. I., Alencar, L. S., and Kanczuk, F. (2006). Demand for Bank Services and Market Power in Brazilian Banking. Working Papers Series 107, Central Bank of Brazil, Research Department.

Ornelas, J. R. H., da Silva, M. S., and Van Doornik, B. F. N. (2022). Informational switching costs, bank competition, and the cost of finance. *Journal of Banking Finance*, 138:106408.

Sanches, F., Silva Junior, D., and Srisuma, S. (2018). Banking privatization and market structure in brazil: a dynamic structural analysis. *The RAND Journal of Economics*, 49(4):936–963.

Silva, M. O. and Lucinda, C. R. (2017). Switching costs and the extent of potential competition in brazilian banking. *EconomiA*, 18(1):117–128.

Stiglitz, J. E. and Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American Economic Review*, 71(3):393–410.

9
Appendix

9.1
Descriptive statistics

Figure 9.1: Proportion of personal loans portfolio to total loans

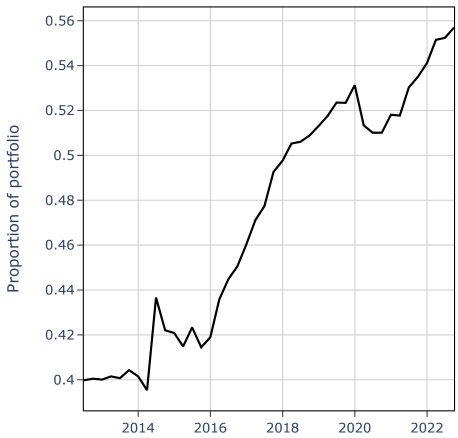


Figure 9.2: Proportion of payroll loans portfolio to personal loans portfolio

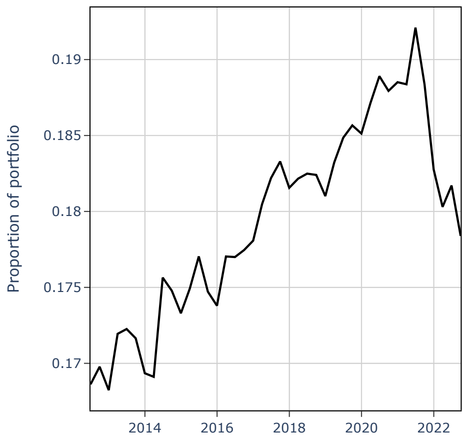


Figure 9.3: Total portfolio for the five largest institutions

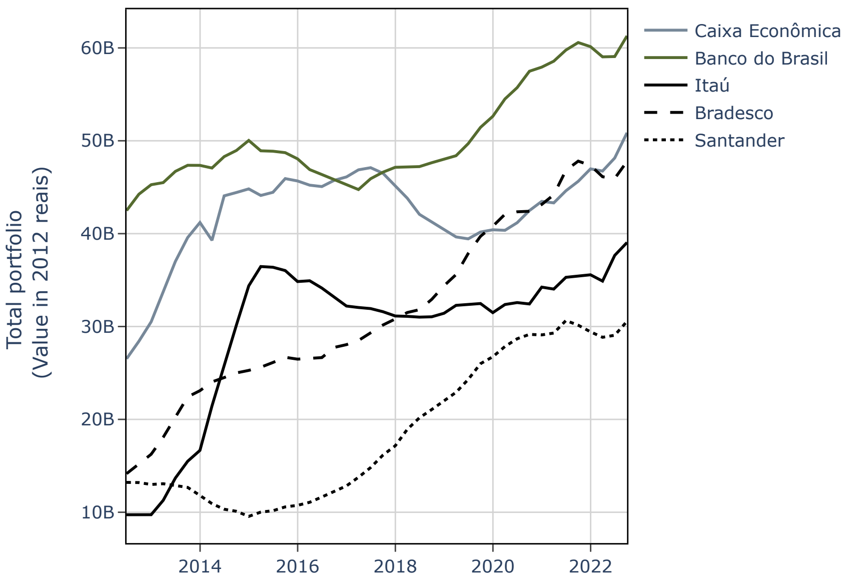
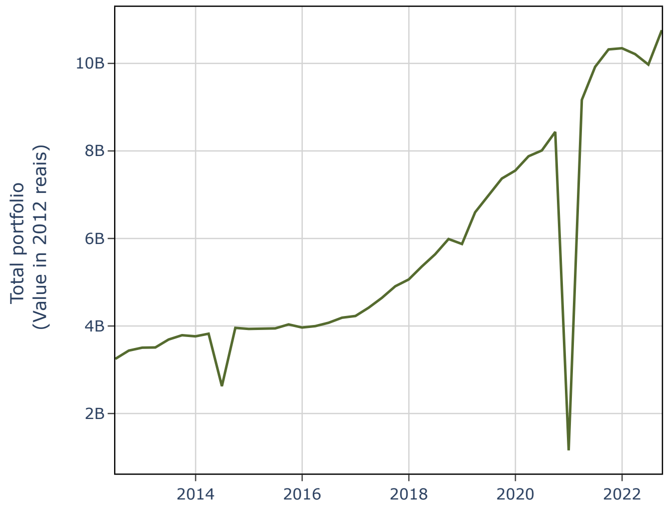


Figure 9.4: Total portfolio for credit cooperatives



9.2 Model

Assume a simplified utility equation of consumer n that chooses to borrow from financial institution j. The utility is composed by two component: V represents characteristics observable by the research, whereas ϵ captures factors that affect the consumer, but are not included in V.

$$U_{nj} = V_{nj} + \epsilon_{nj} \tag{9-1}$$

Since it is not observable, the term ϵ_n is treated as random, following an Extreme Type I distribution f(ϵ_n):

$$f\left(\varepsilon_{nj}\right) = e^{-\varepsilon_{nj}} e^{-e^{-\varepsilon_{nj}}} \tag{9-2}$$

The probability that an option i is chosen in the market is given by the probability that the utility brought by this option is higher than all available options, therefore it is the maximum utility in the set of options:

$$\begin{aligned} P_{ni} &= \operatorname{Prob}\left(V_{ni} + \varepsilon_{ni} > V_{nj} + \varepsilon_{nj} \forall j \neq i\right) \\ &= \operatorname{Prob}\left(\varepsilon_{nj} < \varepsilon_{ni} + V_{ni} - V_{nj} \forall j \neq i\right) \end{aligned} \tag{9-3}$$

The probability can be written as a product of probabilities given by ε_{ni}:

$$P_{ni} \mid \varepsilon_{ni} = \prod_{j \neq i} e^{-e^{-\left(\varepsilon_{ni} + V_{ni} - V_{nj}\right)}}.$$

When integrating over the distribution of ε_{ni}, we have:

$$P_{ni} = \int \left(\prod_{j \neq i} e^{-e^{-\left(\varepsilon_{ni} + V_{ni} - V_{nj}\right)}} \right) e^{-\varepsilon_{ni}} e^{-e^{-\varepsilon_{ni}}} d\varepsilon_{ni}.$$

Which can calculated analytically as:

$$P_{ni} = \frac{e^{\beta' x_{ni}}}{\sum_j e^{\beta' x_{nj}}}$$

9.3
Econometric Specification

9.3.1
First step

Table 9.1: First step: Relevance of instruments

	Dependent variable: Interest rates				
	(1)	(2)	(3)	(4)	(5)
Selic	-0.01***				-0.01***
	(0.00)				(0.00)
Interest rates (firms)		0.05***			0.03*
		(0.02)			(0.02)
Deposits			-3.62***		-4.32***
			(1.06)		(1.05)
Deposits * Selic			0.17		0.23*
			(0.12)		(0.12)
Interest rates (lag)				0.66***	
				(0.02)	
Age	-0.00	-0.00	-0.00	-0.02	-0.00
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Branches	0.10***	0.11***	0.11***	0.04**	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Intercept	-0.06	-0.24	-0.12	-0.94	-0.22
	(1.00)	(1.01)	(1.01)	(0.74)	(0.99)
N	962	962	962	952	962
F-stat	49.44	48.22	47.00	109.41	47.45
Adj. R^2	0.67	0.67	0.67	0.82	0.68

Table 9.2: First step: Relevance of instruments (excluding Caixa)

	Dependent variable: Interest rates				
	(1)	(2)	(3)	(4)	(5)
Selic	-0.01***				-0.01***
	(0.00)				(0.00)
Interest rates (firms)		0.06***			0.03*
		(0.01)			(0.02)
Deposits			-3.29***		-4.18***
			(1.04)		(1.03)
Deposits * Selic			0.11		0.18
			(0.12)		(0.11)
Interest rates (lag)				0.69***	
				(0.02)	
Age	-0.01	-0.01	-0.01	-0.02	-0.01
	(0.02)	(0.02)	(0.02)	(0.01)	(0.02)
Branches	0.11***	0.11***	0.12***	0.04**	0.10***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Intercept	0.47	0.43	0.53	-1.23	0.28
	(1.50)	(1.52)	(1.53)	(1.10)	(1.49)
N	926	926	926	916	926
F-stat	52.98	50.99	49.45	120.53	50.84
Adj. R^2	0.69	0.68	0.68	0.84	0.70

In the alternative model, we estimate the specification represented by Equation 4-4. The regression of the banks' market shares on their prices and characteristics is estimated using OLS and IV, as it is done in Table 6.1. The main difference is that Caixa is excluded from the sample of banks, since it presents a different selection parameter, γ_c. Therefore, the sample comprises $J-2$ institutions.

Table 9.3: First step: Multinomial logit model estimation (excluding Caixa)

	OLS	IV	
		Banks' costs	Lagged prices
	(1)	(2)	(3)
Interest rates	-0.50***	-0.92***	-0.75***
	(0.09)	(0.35)	(0.14)
Branches	0.12*	0.19**	0.13*
	(0.07)	(0.09)	(0.08)
Age	-0.01	-0.03	-0.02
	(0.05)	(0.05)	(0.04)
Intercept	-10.00**	-10.77***	-10.79***
	(4.13)	(4.13)	(4.06)
N	926	926	916
F-stat	413.39	40039.06	41802.77
Adj. R^2	0.95	0.95	0.95

The interest rate coefficient is both negative and statistically significant and its magnitude is close to the coefficients observed in the sample that includes Caixa. The primary distinction arises from the magnitude of the estimated intercept, which is significantly larger. This difference is primarily due to Caixa being the omitted bank fixed-effect when estimating the results presented in Table 6.1.

9.3.2
Second step: outside good

As described in the Econometric specification section, we evaluate how our estimates vary according to different sizes of the outside good.

The size of the market share of the outside good is defined as the proportion of consumers that chose not to buy the product studied. In our framework, we use the extensive margin's choice to identify the parameters of interest. Therefore, it is important to define which consumers are considered for the outside option.

There may be two extreme types of consumers that ended up not purchasing the inside goods: those who never considered taking a loan, and those who were close to getting one but decided against it or were denied. If we consider both types of consumers, the size of the outside good would be large and the utility of the inside option would have to be scaled down considerably in order to generate a large proportion of consumers who do not take loans. This mechanically increases the proportion of type θ equal to one consumers in the population or decreases the γ parameter. Uninterested individuals would be considered as risky ones in this framework.

If we define the outside option based on the second type of people, the outside option would be notably smaller, which would lead to a smaller proportion of type 1 individuals in the distribution or a higher γ parameter. Computing this margin in the data, however, is even more imprecise.

In the paper, we will define the size of the outside option as it is commonly done in the literature, considering all individuals with bank accounts, including those not necessarily interested in getting a loan. We proceed to estimate the same model with smaller values of the outside good, focusing on a smaller group of consumers keen on obtaining a loan.

Table 9.4: Sensibility of estimates to the size of the outside option

	Mean outside good:			
	93%	85%	75%	50%
γ	-42.63	-19.40	-6.07	-5.43
μ	0.52	0.52	0.52	0.53
σ	0.39	0.44	0.48	0.49

As the proportion of the average size of the outside option rises, the parameter γ diminishes in magnitude, while the proportion of type 1 consumers grows. The sign of the parameter remains unchanged, albeit the model fit noticeably worsens when the outside option approaches 50%.

9.4 Results

9.4.1 Second step

We hypothesize that governmental intervention led to a difference in Caixa's operations and risk-taking behavior, possibly affecting the profile of borrowers obtaining loans. In order to fully capture the difference in the selection parameter from Caixa, we further flexibilize the model by allowing the parameter γ_c to vary before and after 2014, when the intervention was discontinued. Table 9.5 presents the estimates for the baseline model and the alternative specification.

Table 9.5: Parameter estimates

	Diff. Caixa effect		Diff. Caixa effect for 2013	
	Banks' costs	Lagged prices	Banks' costs	Lagged prices
γ	-7.35	-17.39	-7.38	-14.26
γ_c	-4.19	-5.67	-4.55	-6.25
γ_c^{2013}			-2.80	-3.42
μ	0.53	0.53	0.53	0.53
σ	0.47	0.44	0.48	0.44

The model fit for the market share and default rate from Caixa are presented in the Figures 9.5 and 9.6. The value of γ_c remains nearly constant, implying that the difference in the selection parameter for Caixa might not be attributed only to the intervention, although γ_c^{2013} is lower in magnitude.

Figure 9.5: Demand for loans in Caixa

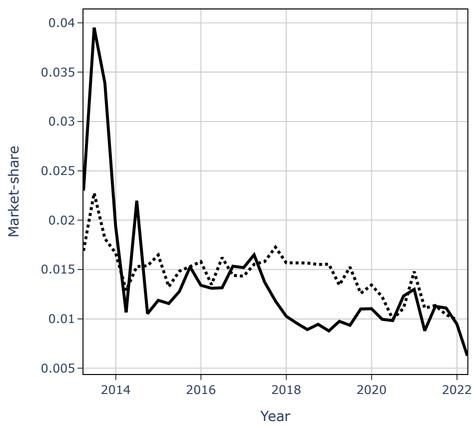


Figure 9.6: Default rate in Caixa