Asymmetric information, search frictions and intermediation in car loans*

Davi Doneda Mittelstadt[1] and Bernardus Van Doornik[2]

[1]University of Wisconsin - Madison
[2]Central Bank of Brazil

January 2026

1 Introduction

In this project, we study how asymmetric information and market power shape the market of auto loans. Cars account for an important share of households' budgets, amounting to around 9% of all household debt in both the United States and Brazil.[1] In this context, search frictions represent an important source of market power, given that prices are personalized and getting quotes might require contacting different lenders. Due to this friction, it is common for auto dealerships to intermediate loans, allowing the consumers to shop for the car and the loan in a single location.

While asymmetric information, search frictions, and intermediation have been studied extensively in the context of credit markets, they tend to be approached in isolation. However, they can interact in important ways: for example, if asymmetric information is severe, we would expect that improving screening technologies, such as credit scores, would generate large gains to consumers. This may not be the case if search frictions are also high, as low-risk consumers might then choose to search less. In this case, part of the welfare gains from better information would be captured by the lenders, instead of flowing entirely to the consumers.

Our goal is to jointly model these important features of auto loans using data from Brazil. This is a setting where informational asymmetries are likely to be high due to the absence of credit scoring, while still possessing a large market for cars and sophisticated lenders. We use data from the Central Bank's Credit Registry, which covers the universe of loans as well as ex-post data on repayments. During our sample period, the Credit Registry was the main source of information on borrowers' creditworthiness for prospective lenders. We reconstruct the information that was available to lenders at the time of loan origination, allowing us to credibly control for all pricing-relevant covariates. One novel aspect of our data is that we also observe which lenders requested data for each borrower in our sample, which we use as a proxy for the search set of each consumer.

*The views expressed in this project are those of the authors and should not be cited to reflect the views of the Central Bank of Brazil.

[1]Sources: SGS/Central Bank of Brazil (2024) and New York Fed Consumer Credit Panel/Equifax (2024).

We first show that asymmetric information is pervasive for both new and used cars. Following Adams, Einav, and Levin (2009), we adapt Chiappori and Salanie (2000) correlation test to the intensive margin, comparing whether individuals who take larger loans are also riskier. We implement this test using random forests, flexibly controlling for the rich set of covariates we observe and that are used by lenders to price the loans. The correlation between residual loan size and residual default risk is positive, which is consistent with riskier individuals self-selecting into larger loans. Estimates are also larger than those found in previous studies, such as Adams, Einav, and Levin (2009) and Crawford, Pavanini, and Schivardi (2018), which is consistent with the absence of credit scores in Brazil during the sample period.

We show that search frictions are also relevant in this market. Using our detailed inquiries dataset, we can directly test if borrowers who search more get better loan terms. A direct regression of loan terms on search intensity shows an unusual positive correlation of interest rates with search intensity, as in Agarwal et al. (2024). To account for the potential endogeneity of search intensity, which might depend on unobserved borrower characteristics, we geocode each borrower's home address and use the distance to physical bank branches as an instrument for search. The assumption is that consumers who live closer to branches have lower search costs. Our IV results show a negative and strong relationship between search intensity and interest rate, pointing to the existence of important search frictions, despite the high degree of loan intermediation through car dealerships.

From these reduced-form results, we build a structural model that incorporates both asymmetric information and search frictions. In the model, individuals can choose how much to search, which loan size to get, and whether to default. They have private information about their willingness to pay for both loan size and risk, allowing riskier individuals to self-select into larger loans. Consumers also have a binary choice of searching with external lenders or accepting the quote offered at the car dealership. One challenge in markets with personalized prices is that researchers observe only transaction prices, not the competing offers. We circumvent this problem following Allen, Clark, and Houde (2019) and Cuesta and Sepulveda (2021), and imposing the conduct assumption that lenders compete as in an English auction. Using standard auction arguments, we can then identify the underlying cost distribution and reconstruct the "missing prices".

Our model contributes to the literature by allowing consumers to choose loan size, which is usually held as fixed in much of the literature, as well as search. We also allow lenders to offer a non-linear pricing menu, based on their expectations that consumers will self-select into different loan sizes.

Our main counterfactual of interest is to estimate the impact of policies that alleviate the information asymmetries. In particular, credit scores were introduced in Brazil in 2020, allowing us to compare our model predictions to the outcomes of the policy. While we do not observe search behavior after 2016 in the data, we are able to estimate a simplified version of the model in the most recent years to quantify how the credit scores affected market outcomes.

This project contributes to a rich literature on auto loans. Closely related are Adams, Einav, and Levin (2009; 2012, 2013), who study asymmetric information in subprime loans in the United States. Our contribution to their work is to consider the joint impact of asymmetric information and search frictions in a setting with multiple competing lenders. Another related contribution is from Grunewald et al. (2023), who analyze inefficiencies introduced when car dealerships intermediate auto loans. While they focus on prime borrowers, for which default risk is negligible, we plan to analyze whether intermediation can also lead to frictions when asymmetric information is a concern.

Another related strand of literature studies market power induced by search frictions in credit markets. An important reference is Allen, Clark, and Houde (2019), who study the relevance of search frictions in the Canadian mortgage market. Much of their modeling approach is followed by Cuesta and Sepulveda (2021), who extend the model to allow for adverse selection in their evaluation of interest rate caps for personal loans in Chile. Given that we are able to observe borrowers' search behavior directly, we do not have to rely as much on modeling assumptions to recover search-related parameters. In this regard, our paper is also close to Agarwal et al. (2024), who also use lenders' inquiries as a measure of search intensity, although in a setting without adverse selection. Also, exploiting discontinuities in the use of credit scores in the United States, Argyle, Nadauld, and Palmer (2023) provide evidence that search frictions cause borrowers to pay higher interest rates and to decrease borrowing amounts in auto loans.

More generally, this project relates to a literature that studies how market power can interact with asymmetric information. The first studies on this originated in the health insurance literature, with Starc (2014) and Mahoney and Weyl (2017) showing that the existence of adverse selection can curb market power. More recently, similar conclusions have been reached by Crawford, Pavanini, and Schivardi (2018) studying credit lines for SMEs in Italy. Our contribution is to extend the study of this interaction to a context with search frictions and intermediation.

2 Institutional background

Credit markets in Brazil are characterized by high default rates, averaging 13% for auto loans between 2016 and 2019. In contrast, in the same period in the United States, the average default was 4.1%.[2] We will argue in this project that they are caused in part by asymmetric information between lenders and borrowers, in which both adverse selection and moral hazard can play a role. Two important institutional features contribute to this market inefficiency. First, there was no widespread credit scoring system in Brazil before 2021, which should exacerbate adverse selection. Before that, lenders would rely mainly on the Central Bank's Credit Registry to evaluate the creditworthiness of loan applicants. Private credit bureaus would inform if a given customer was delinquent at the time of the inquiry, effectively excluding these applicants from formal credit markets, while not informing about the risk of unflagged individuals. Law 12414/2011 allowed credit bureaus to process borrowers' credit history and create credit scores on an opt-in basis, but the take-up rate was low, with less than 5% of the population agreeing to share information. The passing of Law 166/2019 made information sharing the default option, forcing consumers to explicitly opt out of the system if they did not want to share information. After that, data on the majority of the population started being processed by credit bureaus, and informative credit scores were made available to lenders in 2021. Recent evaluations using the staggered rollout of the new scores found that it reduced lenders' markups and increased the supply of credit (Brasil 2021; Hsu Rocha et al. 2024).

Secondly, while collateralized loans offer more security to lenders, the success rate of car repossessions in Brazil is low, around 20%, compared to 70% in the United States (Campos 2023). This should increase the role of moral hazard, when some borrowers might be tempted to default if they can still retain the vehicle. Law 13043/2014, which decreased repossession costs to lenders, might provide useful variation to identify the impact of moral hazard on default rates.[3]

[2] Source: New York Fed Consumer Credit Panel/Equifax.

[3] The law simplified the repossession process by removing notary requirements when notifying the borrower; allowing the car to be collected in a different municipality without requiring official communication between local courts; speeding the process of

Specifically to car loans, a customer has the option of buying a car with an upfront payment, financing, leasing, or group lending (*consórcios*). Their market shares for new cars are 44%, 51%, 5%, and 2%, respectively (ANEF 2019). Conditional on choosing to finance, the typical borrower will take a loan intermediated by the car dealership, which will give a quote from one of the lenders in its network. As we will show later, the networks are small, with a median of 4 lenders for dealers specializing in new cars, and 3 for sellers of used cars.[4] As in the United States, lenders can give financial incentives to dealerships when intermediating a loan (Grunewald et al. 2023). Alternatively, borrowers have the option to search for another lender, which could be their home bank.

Another feature of Brazilian credit markets is the existence of a loan tax.[5] For banks, the rate was set at 3% a year during our sample period (2016-2019).[6] For example, a loan for which a bank would charge 20% would cost 23% a year for the borrower after tax. While loan taxes are not common in the developed world, subsidies can be. For example, since 2025 the United States allows for interest in some auto loans to be deducted from tax payments. How this benefit is passed on to consumers or potentially absorbed by lenders is an empirical question whose answer depends on the economic forces studied in this project.

3 Data

3.1 Data and summary statistics

This project's main dataset is the Credit Registry (SCR) of the Central Bank of Brazil, which stores the universe of loans issued by Brazilian lenders. Our current sample includes all car loans to households originated from 2016 to 2019, as well as ex-post outcomes, including repayment and default, until 2024. This dataset also includes loan characteristics, such as loan size, interest rate, loan term, and collateral (car) value; as well as borrower income. This dataset also includes a dealership identifier when the loan was intermediated. Merging with data from the National Lien System (SNG), we can also observe car characteristics, including make, model, year of manufacturing, average national price, and seller ID. The network of lenders available at every dealership is also observed, given that such intermediaries and their networks have to be registered with the Central Bank.

The main limitations are: 1) we do not observe all characteristics of the car sale, including the down payment and extras. We impute the down payment as the difference between the car value and the loan amount, but we do not observe if the payment was in cash or with a trade-in; 2) we do not observe the rate schedule offered to borrowers, which might include different rates depending on the choice of loan size and term, and potentially requirements such as a minimum down payment; 3) we cannot directly identify customers who shopped for a car loan but did not get one.

The Credit Registry has historically been the main source of borrower information to the lenders, given the absence of credit scores at private credit bureaus. Thus, we can reconstruct the credit history accessed by the lender at the time of origination, containing for each month a snapshot with the number of loans, number

blocking the sale of the vehicle (Melo 2015). Law 14.711/2023 allowed the repossession process not to require court approval, as in the United States, which has the potential to decrease repossession costs significantly. For the analysis of a previous reform that improved repossession rates in Brazil, see Assunção, Benmelech, and Silva (2014).

[4] It is currently unclear if the dealers auction the loan among their network, or if the market is segmented among lenders.

[5] See Brugués and De Simone (2024) for an evaluation of a similar tax in Ecuador.

[6] A flat fee of 0.38% of the loan amount is also charged regardless of loan duration. As the majority of the loans in our sample have durations above 4 years, the flat rate is of lesser importance.

of lender relationships, and debt size for different time windows (e.g.: within the next 30/60/90 days) for all loan types, including mortgages, credit cards, personal loans, and others.[7]

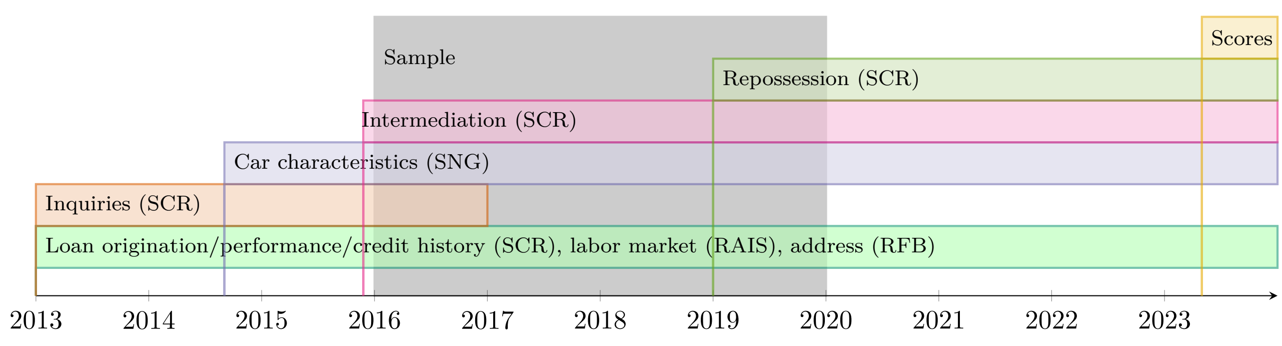
The second dataset is constructed from lenders' inquiries in the Credit Registry. For 2015 and 2016, we observe each credit inquiry, including borrower and lender identifiers, and can merge those into the main dataset. Following Agarwal et al. (2024), we take an inquiry as a measure of search, allowing us to observe how loan characteristics and default vary according to the search intensity.

Lenders can submit inquiries in three modes, which we observe: *web*, when one employee manually logs into the Central Bank's website; *web service*, when the lender requests the data using an API; and *batch*, when the lender requests data of up to 2 million borrowers at the same time. More sophisticated lenders will use *batch* requests to update data on their customers monthly, even if they are not applying for a new loan. This creates a challenge for measuring search, as including all inquiries will overestimate the number of quotes obtained by the borrower, while removing *batch* inquiries will likely underestimate them. Currently, we are defining all inquiries in the 60 days before taking the loan as a quote, but it should be interpreted as an overcount, and will likely mix the effect of search with having an active relationship with many lenders.

In addition, we merge the main dataset with auxiliary data sources: RAIS, an administrative dataset with information on wages, education, type of job, and employer characteristics. This dataset covers the universe of salaried workers in Brazil, comprising 55% of our sample. The remaining borrowers are either self-employed, unemployed, or informal workers. A second dataset is from the Federal Revenue of Brazil, including borrowers' age and address. We geocode each address and calculate the number of branches within radiuses around borrowers.

We also have data on credit scores of each of the five credit bureaus starting from 2023/Q2, when they started being shared with the Central Bank. While not used in this proposal, such data would allow us to gauge our model's predictions against real outcomes from the introduction of credit scores. A summary of all datasets and their periods in which they are available is in Figure 1.

Figure 1: Data availability



In Table 1 we can see summary statistics of our dataset after all merges, pooling across new and user cars unless otherwise noted.[8] First, we see that the average interest rate, 26.2% per year, is much higher than the average funding cost, of 8.7%.[9] We also observe that default rates, defined as a delinquency longer than 90

[7]See Appendix A.4 to a description of this dataset.
[8]For more details on sample selection and some variable definitions, see Appendix A.
[9]We use the future interbank rate as a proxy for funding cost, evaluated at the relevant maturity for each loan.

Table 1: Summary statistics of car loans (2016-2019)

Variable	**Mean**	**SD**	**p0**	**p25**	**p50**	**p75**	**p100**
Loan size (BRL)	26,425.84	16,730.06	4,000.00	15,822.39	22,187.52	32,199.26	237,426.67
Interest Rate (APR)	26.21	11.29	0.00	19.83	22.60	24.82	78.56
Future Interbank Rate (APR)	8.67	2.84	4.36	6.77	7.60	10.79	15.96
Number of installments (months)	41.48	10.95	12.00	36.00	48.00	48.00	60.00
Down payment (%)	36.55	19.56	0.00	22.10	35.09	50.34	89.72
Monthly Income (BRL)	6,800.01	8,419.65	650.00	2,700.00	4,184.89	7,500.00	120,756.53
Relationship (1 = yes)	0.51	0.50	0.00	0.00	1.00	1.00	1.00
Installment value (BRL)	941.09	608.68	35.85	591.44	779.00	1,107.99	69,780.00
PTI (%)	21.90	15.95	0.82	11.81	18.98	27.27	191.08
Previous DTI (%)	49.84	57.46	0.00	11.78	35.03	68.82	866.03
Dealer intermediation (1 = yes)	0.67	0.47	0.00	0.00	1.00	1.00	1.00
Captive lender - new cars (1 = yes)	0.55	0.50	0.00	0.00	1.00	1.00	1.00
New car (1 = yes)	0.21	0.41	0.00	0.00	0.00	0.00	1.00
Car value (BRL)	44,272.68	28,502.37	7,626.33	25,461.00	37,481.60	53,000.00	341,056.00
Default (1 = yes)	0.13	0.34	0.00	0.00	0.00	0.00	1.00
Default 3-months (1 = yes)	0.01	0.08	0.00	0.00	0.00	0.00	1.00
Default 12-months (1 = yes)	0.05	0.22	0.00	0.00	0.00	0.00	1.00
Arrears 15+ days (1 = yes)	0.29	0.45	0.00	0.00	0.00	1.00	1.00
Arrears 15+ days within 1 year (1 = yes)	0.18	0.38	0.00	0.00	0.00	0.00	1.00
# lenders 0-2 km	3.13	2.57	0.00	1.00	3.00	5.00	36.00
# lenders 2-5 km	4.37	4.00	0.00	1.00	4.00	6.00	43.00
# lenders 5-10 km	5.64	6.44	0.00	1.00	5.00	8.00	45.00
Distance nearest lender (km)	1.64	3.01	0.00	0.45	0.91	1.75	248.05
# Inquiries (unique lenders)	2.66	1.27	1.00	2.00	3.00	3.00	19.00
# Inquiries baseline (unique lenders)	1.91	1.09	0.00	1.00	2.00	3.00	17.00
Network size	4.71	4.74	0.00	0.00	4.00	9.00	16.00
# Inquiries out-of-network	1.17	1.58	0.00	0.00	0.00	2.00	15.00
Number of observations	8,523,323						

Note: For reference, the average exchange rate in this period was 3.57 BRL for 1 USD. The relationship dummy denotes borrowers with an existing relationship with the lender for more than 1 year before taking the loan. Payment-to-income (PTI) is the loan-specific monthly payment divided by monthly gross income. Debt-to-income (DTI) is the sum of all debt due within 30 days divided by monthly gross income, calculated the month before taking the auto loan. Network size is imputed as zero when we couldn't match the seller with an intermediary registered with the Central Bank. The number of baseline inquiries was calculated as the number of inquiries between 60 and 90 days before origination of the auto loan. All inquiries and network size variables are only for 2016.

days, are high, averaging 13% over the loans' term, and reaching 5% in the first 12 months. If we adopt a less strict measure of delinquency, including all loans that were in arrears for more than 15 days, the fraction of delinquent loans reaches 29%. Despite some issues with our measure of income, discussed in Appendix A.3, our measure of payments-to-income (PTI) is roughly in line with Argyle, Nadauld, and Palmer (2023), who study car loans in the United States. We can also see that intermediation is widespread in this market, with 67% of all loans being intermediated. For new cars, lenders who are vertically integrated with car manufacturers (e.g., General Motors Bank and Stellantis Bank) are responsible for 55% of originations.

From the inquiries dataset, available only for 2016, we can observe that the median borrower has 3 unique lenders with inquiries in the 60 days before auto loan origination. However, as we discussed before, some of these can be related to previous loans. For this reason, we also calculate a "baseline" number of lenders, measured between 60 and 90 days before loan origination, and presumably not related to the car acquisition. By comparing the means, we can calculate that borrowers have 0.75 more lenders consulting their credit history when they shop for an auto loan. That this number is less than one means that a significant portion of borrowers seek quotes with institutions that were already submitting inquiries. We can also observe that the majority of borrowers do not seek quotes from lenders outside of the dealer's network.

3.2 Asymmetric information

Asymmetric information in this context can be due to adverse selection or moral hazard. Adverse selection can be described as a setting where borrowers differ in risk, and riskier types take larger loans, while lenders do not perfectly screen types. Adverse selection might coexist with moral hazard, in which larger monthly payments make the customer more likely to default. One of the goals of this project is to separately identify each of these frictions, given that the policy recommendations for each are different: credit scoring tends to alleviate adverse selection, while better repayment incentives are used to curb moral hazard (Adams, Einav, and Levin 2009). For the moment, we are abstracting away from this distinction.

Testing for the existence of asymmetric information is not straightforward in credit markets. Fixing contract characteristics, a test in the spirit of Chiappori and Salanie (2000) would compare ex-post outcomes between customers who take a loan and those who do not. In the presence of either adverse selection or moral hazard, we would expect customers who take a loan to be riskier. Such a test is not possible in my setting, given that I do not observe customers who choose not to borrow. Also, default is only defined for borrowers, such that we would need exogenous variation or strong assumptions to compute counterfactual default rates for consumers who did not take a loan.[10]

The alternative is to use the intensive margin of borrowing: conditional on taking a loan, riskier customers borrow more than less risky ones. In auto loans, taking the car choice as exogenous, this would translate to riskier types choosing lower down payments (Adams, Einav, and Levin 2009). Econometrically, we want to test if consumers who choose lower down payments (or, equivalently, larger loan sizes) are more likely to default, after controlling for characteristics that affect pricing. My basic specification for the asymmetric information test estimates two equations:

$$\begin{aligned} \ln(\text{loan size}) &= f(X^L, X^C, X^{CH}, X^B) + \varepsilon \\ \text{default} &= g(X^L, X^C, X^{CH}, X^B) + \eta, \end{aligned} \tag{1}$$

where X^L denote loan characteristics, including number of installments, year/month, dealer ID; X^C includes car brand, model, trim, year, and value; X^L include customer characteristics, including income, region, and demographics; X^{CH} includes the borrower's credit history[11], and X^B include lender ID, conglomerate ID, and a dummy indicating an existent relationship with the borrower. We use random forests to approximate $f(\cdot)$ and $g(\cdot)$, and calculate $cor(\varepsilon, \eta)$. Table 2 contains measures of fit for each model. The correlations were calculated considering the "extended" specification, pooling new and used cars.

Correlation results are available in Table 3. For both new and used cars the correlation is positive, at 0.057 for new cars and 0.068 for used cars. They are also relatively high compared to the literature: in Crawford, Pavanini, and Schivardi (2018) the comparable estimate is 0.03, and in Einav, Jenkins, and Levin (2012) it is 0.02. The estimates are robust to excluding customers with an existing relationship with lenders, which we defined as existing for more than one year before loan origination. We can also measure that intermediated loans show a higher correlation between loan size and default compared to non-intermediated loans. We note

[10]For example, Crawford, Pavanini, and Schivardi (2018) assume a Heckman selection model to their extensive margin test. This assumption is also used in the structural models of Einav, Jenkins, and Levin (2012) and Cuesta and Sepulveda (2021).

[11]See Appendix A.4 for more details on the available variables

Table 2: Random forests fit measures

Statistic	Baseline	Extended	Full
Default			
Brier score	0.095	0.092	0.091
ROC AUC	0.753	0.788	0.795
Loan size (log)			
R-squared	-	0.784	-
RMSE	-	0.243	-
Interest rate			
R-squared	-	0.772	0.784
RMSE	-	4.040	3.930

Note:
The baseline model contains contains only preexisting covariates, including income, credit history, and region. The extended model includes all the variables from the baseline model plus the identity of lender, seller, and car. The full model adds the down payment choice.

that these results are not causal: it might be that the correlation would have been even higher if the loans were not intermediated. We also measured that loans originated by captive lenders (vertically integrated with car manufacturers) show a slightly higher correlation than those established with conventional lenders when we focus on new cars. For used cars, captives' share is below 5%, so the results are less interesting. When we focus on used cars, in which the fraction of P2P sales is 71%, we see that those loans present a slightly lower correlation than the average. Finally, for a sanity check in our results, we can see that the correlation is increasing in baseline risk, which is calculated using only preexisting borrower characteristics (as in a FICO score). That is, unobservable risk is increasing in observable risk.

3.3 Nonlinear pricing

Another feature of our setting is nonlinear pricing: lenders tend to offer lower rates for consumers who take a smaller loan. We can explain this through two channels: the first purpose is, as in Mussa and Rosen (1978), to charge more from consumers with a higher willingness to pay for larger quantities. The second is to better screen consumers, as, in a context with adverse selection, borrowers who choose larger loans are also more likely to default. Given our focus on adverse selection, we are considering only the second channel. One challenge is to identify this from data, as we do not observe the rate schedule that is offered to consumers, only the combination of loan size and rate that was picked.

We will show that rates are increasing in loan size with two approaches: first, we adapt the correlation test from Chiappori and Salanie (2000), but instead of calculating the residual correlation between loan size and default, we estimate the residual correlation between rates and loan size. In the second, we directly regress rates on down payment in an attempt to recover the rate schedule from the transaction data. To flexibly control for all other relevant variables, we perform both procedures using random forests, with the same controls as in Section 3.2. One implicit assumption in both procedures is that we have sufficient variation in the choice of loan size to identify the relationship with rates. This will not be the case if, for example, lenders set strict down payment requirements and the totality of borrowers choose the minimum.

In Table 4 we can see the results of the correlation test. The correlation is very significant for new cars, and

Table 3: Correlation test

Sample	New cars	Used cars
All	0.057	0.068
Relationship		
New customers	0.067	0.077
Existing customers	0.044	0.056
Intermediation		
Intermediated	0.062	0.073
Non-intermediated	0.045	0.056
Captive lender		
Captive	0.062	0.103
Non-captive	0.054	0.066
Seller type		
B2P (dealership)	0.057	0.078
P2P	-	0.064
Baseline risk quartile		
1	0.031	0.059
2	0.050	0.071
3	0.068	0.073
4	0.086	0.080

lower but still positive for used cars. The result for used cars can be due to the absence of variation in the data.

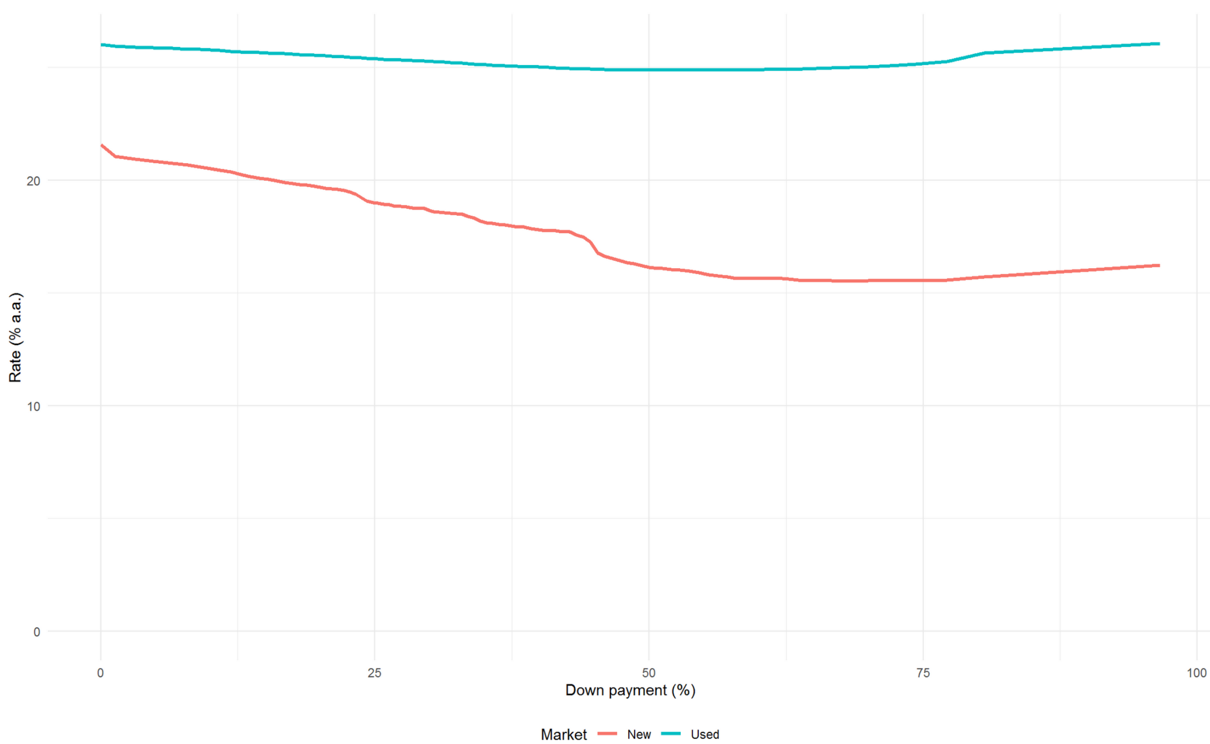
In Figure 2 we can see the result of the direct regression of rates on down payment fraction. For calculating the partial dependence plot, we estimate random forests on our sample using cross-validation. After estimation, we use the model to predict out-of-sample counterfactual rates for different down payment levels. In practice, we do this procedure for a sample of 1,000 observations and average the results. The results confirm that rates are lower for larger down payments, but this decline is steeper for new cars when compared to used cars.[12]

[12]The estimates when the down payment is higher than 60% should not be taken at face value, as fewer than 10% of the observations are in this range.

Table 4: Residual correlation between loan size and rates

Sample	New cars	Used cars
All	0.147	0.039
Relationship		
New customers	0.148	0.047
Existing customers	0.149	0.030
Intermediation		
Intermediated	0.165	0.049
Non-intermediated	0.090	0.018
Captive lender		
Captive	0.136	0.162
Non-captive	0.182	0.033
Seller type		
B2P (dealership)	0.147	0.084
P2P	-	0.023
Baseline risk quartile		
1	0.115	0.086
2	0.148	0.060
3	0.161	0.023
4	0.162	-0.018

Figure 2: Partial dependence plot of rates on down payment (%)



3.4 Search costs

One characteristic of markets with personalized pricing is the existence of search frictions, given that potential borrowers have to actively seek quotes from lenders. In most settings, it is not straightforward to measure such inefficiencies because search is not observed. In our setting, we can leverage our dataset with inquiries

Table 5: Regression of loan outcomes on search intensity, using OLS and 2SLS

	Interest rate	Down payment (%)	Default	Interest rate	Down payment (%)	Default
		OLS			IV	
	(1)	(2)	(3)	(4)	(5)	(6)
# inquiries	0.216***	0.026	0.004***	-3.47***	3.15*	-0.060**
	(0.013)	(0.021)	(0.0004)	(0.639)	(1.86)	(0.024)
Observations	1,405,776	1,405,776	1,405,776	1,405,776	1,405,776	1,405,776
R^2	0.648	0.462	0.124	0.537	0.444	0.100
RMSE	5.28	13.9	0.305	6.06	14.1	0.309
Dependent variable mean	25.6	38.0	0.121	25.6	38.0	0.121
F-test (1st stage), # inquiries				21.6	21.6	21.6

Table 6: Regression of loan outcomes on search intensity by market, using OLS and 2SLS

	Interest rate	Down payment (%)	Default	Interest rate	Down payment (%)	Default	Interest rate	Down payment (%)	Default	Interest rate	Down payment (%)	Default
			OLS						IV			
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
# inquiries	0.193***	-0.663***	0.006***	0.159***	0.215***	0.003***	-1.56*	1.14	0.001	-2.80***	3.49***	-0.065**
	(0.014)	(0.050)	(0.0005)	(0.013)	(0.025)	(0.0006)	(0.917)	(1.55)	(0.034)	(0.446)	(1.05)	(0.026)
Observations	371,123	371,123	371,123	1,034,552	1,034,552	1,034,552	371,123	371,123	371,123	1,034,552	1,034,552	1,034,552
R^2	0.572	0.550	0.119	0.534	0.396	0.111	0.549	0.545	0.119	0.401	0.374	0.087
RMSE	5.79	13.6	0.201	4.49	13.8	0.334	5.95	13.6	0.201	5.09	14.1	0.338
Dependent variable mean	17.0	45.1	0.048	28.7	35.5	0.147	17.0	45.1	0.048	28.7	35.5	0.147
F-test (1st stage), # inquiries							5.42	5.42	5.42	19.6	19.6	19.6

to directly measure if consumers with inquiries from more lenders obtain better terms.

First, we estimate a linear model with default and interest rates as dependent variables and the number of inquiries from distinct lenders as the covariate of interest, as a proxy for search intensity. We add as controls borrowers' characteristics, credit history, car characteristics, and lender identity. The results of the estimation with OLS are in columns 1-4 of Table 5. Interestingly, the results show a positive correlation between both rates and default with search, which is inconsistent with the canonical search models. One explanation is that consumers who are riskier are more likely to be rejected, as in Agarwal et al. (2024).

In the second approach, we use instrumental variable to control for the fact that search intensity is endogenous. We instrument the number of inquiries with the proximity to branches. To do so, we geocode each borrower in our sample and also each active bank branch. We compute the number of distinct branches and lenders within 2km, 2-5km, 5-10km, and 10-20km, as well as the distance to the nearest branch[13]. The idea is that consumers who are exogenously located close to physical branches have lower search costs, as in Argyle, Nadauld, and Palmer (2023). Results are available in columns 5-8 of Table 5. In the current specification, the first stage is not high enough for any meaningful inference for new cars, but it is high enough for used cars. Results show that each additional quote reduces the interest rate for used cars by 2.1 p.p., or by 7% of the average rate, or 0.21 standard deviations. It is important to stress that the majority of the variation is coming from one additional quote, such that we would expect the results to decrease for further quotes.

4 Model

The model has three main components: 1) asymmetric information; 2) search costs; 3) nonlinear pricing.

The timeline is as follows:

[13]Given the issues in making inference based on 2SLS when there with multiple instruments, we intend to include the Anderson-Rubin test and include an alternative estimation with LIML.

0. Consumers choose an auto dealership and a car outside of the model
1. Borrowers choose the search intensity for a car loan, which determines their consideration set. If they do not search, they get a free quote at the dealer, including only lenders in the dealer's network. If they search, they can access external lenders not available at the dealer
2. Borrowers choose a loan size, taking expectations over the rate for that option, conditional on their consideration set
3. Borrowers choose the lender from which they borrow, conditional on the consideration set and loan size
4. Borrowers choose to default or not

We can include consumers' observables in the model, but we will omit it for simplicity. Thus, we can consider this a model for consumers with the same observed characteristics and the same exogenous loan characteristics (e.g., term and car value).

4.1 Demand

Borrower i has multidimensional private information $\theta_i = \{\theta_i^q, \theta_i^d\}$ that affects their willingness to pay for larger loans and their risk of default, respectively, $\theta_i \sim F$. We will present the model backwards.

4.1.1 Default

After borrowing, the borrower's default shock θ_i^d is realized, which will determine whether they default through a function $g(\cdot)$. For simplicity, we assume that consumers who default do so in the first month, so that they do not repay any fraction of the loan.

$$D = 1(g(\theta_i^d) > 0) \tag{2}$$

The effect of θ on default will be related to adverse selection (in the intensive margin) as long as θ^d is correlated with θ^q. In this case, consumers who are more likely to take larger loans are also more likely to default.

4.1.2 Lender choice

Conditional on loan size q and consideration set S, consumers choose the lender j that offers the lowest interest rate. That is, there is no lender differentiation, as in Allen, Clark, and Houde (2019) and Cuesta and Sepulveda (2021). There is no outside option.

4.1.3 Loan size

In the second stage, consumers choose a loan size q over a discrete menu to maximize their utility. Consumers form expectations over the interest rate p they will face in the next stage for each level of q, conditional on their choice of consideration set S. We have random coefficients on q that depend on the consumer's private information θ_i^q. There is an idiosyncratic shock ε_{iq}^q.

$$U^Q = \alpha_i^Q q - E(p|q,S) + \varepsilon_{iq}^Q, \quad \alpha_i^Q = h(\theta_i^q) \tag{3}$$

To improve interpretability, we can think of $q = m \cdot p^C$, where m is the down payment expressed as a fraction of the loan size, and p^C is the exogenous car price. Then, we could translate the choice of loan size, which is continuous and fundamentally dependent on the car price, to a percentage point scale of m, which is common across all contracts.

4.1.4 Search

Consumers choose nonsequentially how much to search, and that determines their consideration set $S \in \mathcal{S} \equiv \{S_1, S_2\}$. They can choose among two options: search only at the dealer (S_1) at zero cost, as they already visited the dealer for choosing the car, which happens outside of the model; or search additionally with an external lender not available at the dealership (S_2), incurring search cost κ. This choice can depend on the availability of branches close to their address, existing relationships, or other observable characteristics (omitted in the notation below). Their indirect utility over search can be expressed as

$$U^S = E(U^Q|S) - \kappa \times 1\{search\} + \varepsilon_{is}^S \tag{4}$$

Where $E(U^Q|S)$ is the inclusive value of choosing the consideration set S. For simplicity, we can think that S_1 includes all lenders in the dealer's network (observed), and S_2 is the borrower's home bank. The model can be extended to allow for searching at multiple outside lenders. For example, we could assume a constant search cost κ and endogenize the number of external lenders n, $n \in N$, where N is the total number of lenders available in the borrower's i market.

Another possible extension is due to the fact that we do not perfectly observe the consideration sets, given that lenders may have submitted inquiries on the borrower due to reasons other than a request for a car loan quote. For this reason, we might model that borrowers draw random subsets from the available lenders, as in Goeree (2008) and Cuesta and Sepulveda (2021).

4.2 Supply

Lenders compete in rates as in an English auction, including all lenders in the borrower's consideration set. Lenders observe the consumer's loan size and search choice. Their profits for a given loan are given by

$$\Pi_j = [1 - D(\theta_i^d)] \times p \times q - c_{ij} \times q, \tag{5}$$

where $D(\theta_i^d)$ is the expected default rate of consumer i; p denotes the interest rate; q is the loan size; and c_{ij} denotes the marginal cost of lender j providing a loan to customer i, which can be interpreted as a match value, as in Allen, Clark, and Houde (2019) and Cuesta and Sepulveda (2021). This match value can be due to the perceived value of establishing a relationship and the opportunities for cross-selling products.

4.3 Equilibrium

Given the English auction assumption, it is optimal for lenders to bid their "valuations" in stage 3, conditional on the choice of loan size (which is informative of types) and choice set. In equilibrium, the lender with the lowest cost wins and charges the rate implied by the lender with the second lowest cost. The equilibrium price solves:

$$E_\theta[\Pi] = [1 - E[D(\theta_i^D)|q, S]] \times q \times p^*(q, S) - c_{i(2)} \times q = 0 \tag{6}$$

Which can be expressed as:

$$p^*(q,S) = \frac{c_{i(2)}}{1 - E[D(\theta_i^D)|q,S]} \tag{7}$$

An equilibrium is defined as:

- Given expected default rates for each $q \in Q$ and $s \in S$, lenders submit rates $P_j(q,S)$ to maximize their profits, competing as in an English auction
- Observing rates $P_j(q,s), j \in s$, where q is the previous choice of loan size and s the chosen consideration set, consumers maximize their utility by choosing the lender that offers the lowest rate
- Given the expected menu of interest rates $P_j(Q,s), j \in s$, consumers choose loan size $q \in Q$
- Given the expected menu of interest rates $P_j(Q,S), j \in S$, and expected loan size choice $q \in Q$, consumers optimally choose whether to search
- Lenders' and borrowers' expectations are consistent

4.4 Sketch of algorithm for solving model

0. Fix all parameters in the model
1. Guess a distribution of types over loan size and a distribution of types over consideration sets
2. Calculate expected default rates for each loan size bin and consideration set
3. Given expected default, calculate expected prices for each loan size bin and consideration set using Equation 7
4. Given expected prices, calculate the probability of choosing a loan size for each consideration set and θ
5. Given expected prices and loan size, calculate the probability of choosing consideration sets for θ
6. Compare the distribution of types over q and S. If different from the guess, update and rerun until convergence

4.5 Econometric model

The model's primitives of interest are:

- Search costs κ
- Distribution of lenders' costs
- Distribution of types (in particular, the correlation between θ^Q and θ^D, which denotes the level of adverse selection)
- Default function
- Utility for loan size

We will make parametric assumptions that will give us tractability for solving the model. We will assume that $\varepsilon_{iq}^Q, \varepsilon_{is}^S$ $T1EV$, as standard in the discrete choice literature, as this gives us the usual closed-form expression for the probability of each choice. We will also follow Einav, Jenkins, and Levin (2012), Crawford, Pavanini, and Schivardi (2018), and Cuesta and Sepulveda (2021) and assume that the private types θ $N(\mu, \Sigma)$, with correlation parameter ρ. Finally, we assume that functions $g(\cdot)$ and $h(\cdot)$ in Equation 9 and Equation 3 are linear. Then, we can rewrite the default and loan size equations as follows:

$$
\begin{aligned}
D &= 1(\theta_i^d > 0) \\
U^Q &= \theta_i^q q - E(p|q, S) + \varepsilon_{iq}^Q
\end{aligned}
\tag{8}
$$

Given these assumptions, we can calculate the expected default rate $E(D|Q = q, S = s)$:

$$
\begin{aligned}
E(\theta^d > 0|Q = q, S = s) &= \int E(\theta^d > 0|Q = q, S = s, \theta^q) dF(\theta^q) \\
&= \int E(\theta^d > 0|\theta^q) P(Q = q|S = s, \theta_q) P(S = s|\theta_q) dF(\theta^q),
\end{aligned}
$$

where

$$
\begin{aligned}
f(\theta^d|\theta^q) &= N\left(\mu^d + \rho\frac{\sigma^d}{\sigma^q}(\theta^q - \mu^q), \sigma^d\sqrt{1-\rho^2}\right) \\
P(Q = q|S = s, \theta_q) &= \frac{exp\{\theta^q q - E(p|q, s)\}}{\sum_k exp\{\theta^q k - E(p|k, s)\}} \\
P(S = s|\theta_q) &= \frac{exp\{E(U^Q|S = s) - \kappa \times 1\{search\}\}}{exp\{E(U^Q|S = 0)\} + exp\{E(U^Q|S = 1) - \kappa \times 1\{search\}\}} \\
f(\theta^q) &= N(\mu^q, \sigma^q)
\end{aligned}
$$

4.6 Comments

A major obstacle to empirically studying markets with personalized pricing is that the econometrician usually does not observe all quotes available to the borrower, only the chosen contract. Our strategy is to set a supply-side restriction, assuming that lenders compete as in an English auction, and leveraging it to identify the lenders' cost distribution, as in Allen, Clark, and Houde (2019) and Cuesta and Sepulveda (2021). Another alternative that relies on supply-side restrictions is given by D'Haultfœuille, Durrmeyer, and Février (2019), but their approach focuses on price discrimination based on observable characteristics, which is less suited for our analysis. One alternative that has been used in credit markets is to use instead the usual differentiated products framework of Berry, Levinsohn, and Pakes (1995) and impute the missing prices, as in Crawford, Pavanini, and Schivardi (2018) and Brugués and De Simone (2024). However, there is significant residual dispersion in prices after controlling for the available covariates, despite the richness of our controls and the flexible estimation, as shown in Table 2. Therefore, we think this approach is not well-suited for our application. A different approach is given by Cosconati et al. (2025), who design an outer fixed point that accounts for the missing prices and can be used within the differentiated products. Our current choice is mainly due to tractability, but an in-depth analysis of the gains in adopting their approach is in the roadmap for this project.

We make different modelling assumptions compared to the most recent strategies implemented to estimate nonlinear pricing. A leading example is given by Luo, Perrigne, and Vuong (2018), who exploit the firm's first order condition to show that the model's primitives are identified nonparametrically. Separately, D'Haultfœuille and Février (2020) use exogenous variation in the principal's rate schedule to recover non-parametrically the agents' type distribution and parameters of their cost function. Both approaches can be combined, as shown in Kang and Silveira (2021). A new non-parametric approach to identify a multidimensional type distribution was introduced by Aryal et al. (2025), which does not rely on optimality conditions and therefore allows for flexible forms of competition. Translating their framework from insurance to credit markets is not straightforward, as they require an instrument to identify what, in our context, would be the preference for loan size, for which we have no candidates.

Our departure is due to extra features contained in our analysis that complicate a direct application: the most important is that there is no monopoly, as there are many lenders competing for borrowers. Also, we do not observe the rate schedule posted by lenders, only the chosen contract, which poses an extra obstacle to identification.

As surveyed by Stole (2007), there are theoretical models for second-degree price discrimination that relax the monopoly assumption, but those models quickly lose tractability for an empirical analysis. Closer to our analysis is Lester et al. (2019), who develop a general model that could be applied to credit markets including adverse selection and search frictions, while allowing lenders to set rate/quantity schedules. However, their analysis is fully theoretical.

The empirical papers that are closest to our analysis in taking contract choice by lenders explicitly are Adams, Einav, and Levin (2009) and especially Einav, Jenkins, and Levin (2012), who endogenize the minimum down payment required from borrowers. Our model nests this decision, as it allows lenders to set an infinite interest rate for loan sizes in which only high-risk borrowers would participate.

As previously mentioned, in our empirical setting, we observe a positive correlation between search intensity and interest rates, as in Agarwal et al. (2024). While they model this approach as being due to a positive probability of rejection, we instead model lender pricing explicitly, allowing borrowers to search not only for approvals but also for better quotes.

4.7 Counterfactuals

We have three counterfactuals of interest: 1) a decrease in the level of adverse selection due to an improvement in credit scores; 2) pass-through of loan taxes/subsidies; 3) a decrease in the costs of repossession. We now cover how each of them would be performed given our current model.

The introduction of credit scores would map as a reduction in the correlation between willingness to pay for larger loans and the risk of default, as in Crawford, Pavanini, and Schivardi (2018). This would decrease rates for low-risk consumers, potentially enabling them to choose larger loans. On the other hand, if consumers choose to search in fewer lenders, this might lead to more market power, granting part of the surplus to lenders. This would be an endogenous increase in market power. Quantifying these effects would be one contribution from this project.

To run the second counterfactual, we would need to include taxes in the model, which is straightforward. Note that, without choice on loan size and search, there will always be either full pass-through, in the case when there is no adverse selection and therefore no changes in default rates (as in Allen, Clark, and Houde 2019), or overshifting, when default increases (as in Cuesta and Sepulveda 2021). When we endogenize loan size and search, consumers can choose to take a smaller loan with potentially lower rates, which would allow for tax undershifting.

The third counterfactual would require more additions to the model. We would have to microfound repayment to include a cost to default, which depends on the probability of the lender repossessing the car. Alternatively, we could extend the supply side, adding a fixed cost for lenders for repossessing a car and a scrap value. In any case, this would require some data cleaning: our dataset provides information on car repossession since 2019 and on sales of loans to other lenders and non-financial firms, but those are not readily available at this point.

5 Next steps

In the model, there are some absent features that might be relevant to include. First, the current model reflects only adverse selection, not moral hazard. Given that Adams, Einav, and Levin (2009) points to its relevance for auto loans in the United States, it might be relevant to include it, which would require changing Equation 9 to depend directly on loan size q and interest rate p:

$$D = 1(g(\theta_i^d, q, p) > 0) \tag{9}$$

This would require another fixed point to solve the model, which might complicate the algorithm. Also, it is unclear if we have sufficient exogenous variation in the data to separately identify adverse selection and moral hazard without relying only on parametric identification.

Another extension would be to endogenize the loan term. While it has been standard in the literature to take both loan size and term as exogenous, once we endogenize loan size it might be too restrictive to hold the term fixed: for example, if a borrower changes the down payment from 50% to zero, the increase in monthly payments may not fit her budget, requiring her to also increment the number of installments. One way to implement this without changing much of the model would be to consider a joint choice of loan size and term, in which only the alternatives that are below an income threshold are available. This would add an extra dimension for the lenders to screen consumers.

One limitation of the current setup is that in most situations it is profitable for lenders to offer a loan, as they can always increase prices to account for higher default rates. The exception is when larger loan sizes unravel due to the adverse selection, but in simulations, it was hard to reproduce this behavior. To endogenize more easily such down payment requirements, there are three non-mutually exclusive alternatives that could be added: 1) including moral hazard, as described before, which would affect default rates also for inframarginal consumers; 2) adding fixed costs of default to the lenders, which would make some loans non-profitable; 3) increasing adverse selection in the extensive margin by adding an outside choice, which could be interpreted as buying the car without a loan, or not buying the car at all. The restriction on adding the outside option is data, as we do not have a clear measure of the size of the outside option. One avenue we are pursuing is to gather data on all inquiries submitted by vertically integrated lenders (e.g., General Motors Bank), and consider the share of consumers who get an inquiry but don't take a loan as the share of the extensive margin. One limitation is that it would be available only for new cars, as such lenders are not relevant in the used car market. The use of other lenders is not possible, given that they offer other financial products, and inquiries may not be related to auto loans.

Other than model extensions, one necessary step is to think carefully about parameter identification. This is especially important given that estimation will be performed using method of moments, and the parameters must have a clear map to the moments we choose for the estimation.

6 References

Adams, William, Liran Einav, and Jonathan Levin. 2009. "Liquidity Constraints and Imperfect Information in Subprime Lending." *American Economic Review* 99 (1): 49–84.

Agarwal, Sumit, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao. 2024. "Searching for Approval." *Econometrica* 92 (4): 1195–1231. https://doi.org/10.3982/ECTA18554.

Allen, Jason, Robert Clark, and Jean-François Houde. 2019. "Search Frictions and Market Power in Negotiated-Price Markets." *Journal of Political Economy* 127 (4): 1550–98.

ANEF. 2019. "Boletim Anual." http://anef.com.br/administrador/files/pdfs/7cd8204e014e764f658760f0dc138cb3.pdf.

Argyle, Bronson, Taylor Nadauld, and Christopher Palmer. 2023. "Real Effects of Search Frictions in Consumer Credit Markets." *The Review of Financial Studies* 36 (7): 2685–2720. https://doi.org/10.1093/rfs/hhac088.

Aryal, Gaurab, Isabelle Perrigne, Quang Vuong, and Haiqing Xu. 2025. "Econometrics of Insurance with Multidimensional Types." *Quantitative Economics* 16 (1): 267–94. https://doi.org/10.3982/QE1071.

Assunção, Juliano J., Efraim Benmelech, and Fernando S. S. Silva. 2014. "Repossession and the Democratization of Credit." *The Review of Financial Studies* 27 (9): 2661–89. https://doi.org/10.1093/rfs/hht080.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile Prices in Market Equilibrium." *Econometrica* 63 (4): 841–90. https://doi.org/10.2307/2171802.

Brasil, Banco Central do. 2021. "Análise Dos Efeitos Do Cadastro Positivo," April.

Brugués, Felipe, and Rebecca De Simone. 2024. "Taxation When Markets Are Not Competitive: Evidence from a Loan Tax," January.

Campos, Álvaro. 2023. "Febraban Mira Veto à Busca e Apreensão de Veículos." *Valor Econômico*, November. https://valor.globo.com/financas/noticia/2023/11/13/febraban-vai-trabalhar-junto-ao-congresso-para-derrubar-veto-a-busca-e-apreensao-de-veiculos.ghtml.

Chiappori, Pierre-Andre, and Bernard Salanie. 2000. "Testing for Asymmetric Information in Insurance Markets." *Journal of Political Economy* 108 (1): 56–78. https://doi.org/10.1086/262111.

Cosconati, Marco, Yi Xin, Fan Wu, and Yizhou Jin. 2025. "Competing Under Information Heterogeneity: Evidence from Auto Insurance."

Crawford, Gregory S., Nicola Pavanini, and Fabiano Schivardi. 2018. "Asymmetric Information and Imperfect Competition in Lending Markets." *American Economic Review* 108 (7): 1659–1701.

Cuesta, José Ignacio, and Alberto Sepulveda. 2021. "Price Regulation in Credit Markets: A Trade-Off Between Consumer Protection and Credit Access." Working paper. Rochester, NY.

D'Haultfœuille, Xavier, Isis Durrmeyer, and Philippe Février. 2019. "Automobile Prices in Market Equilibrium with Unobserved Price Discrimination." *The Review of Economic Studies* 86 (5): 1973–98. https://doi.org/10.1093/restud/rdy064.

D'Haultfœuille, Xavier, and Philippe Février. 2020. "The Provision of Wage Incentives: A Structural Estimation Using Contracts Variation." *Quantitative Economics* 11 (1): 349–97. https://doi.org/10.3982/QE597.

Einav, Liran, Mark Jenkins, and Jonathan Levin. 2012. "Contract Pricing in Consumer Credit Markets." *Econometrica* 80 (4): 1387–1432.

———. 2013. "The Impact of Credit Scoring on Consumer Lending." *The RAND Journal of Economics* 44 (2): 249–74. https://doi.org/10.1111/1756-2171.12019.

Goeree, Michelle Sovinsky. 2008. "Limited Information and Advertising in the U.S. Personal Computer

Industry." *Econometrica* 76 (5): 1017–74. https://doi.org/10.3982/ECTA4158.

Grunewald, Andreas, Jonathan Lanning, David Low, and Tobias Salz. 2023. "Auto Dealer Loan Intermediation: Consumer Behavior and Competitive Effects."

Hsu Rocha, Roberto, Javier Feinmann, Mariana Mercucci, and Josival Leite. 2024. "The Effects of Revealing Borrowers' Information on Credit Allocation, Defaults and Entrepreneurship."

Joaquim, Gustavo, Bernardus van Doornik, and José Renato Haas Ornelas. 2023. "Bank Competition, Cost of Credit and Economic Activity: Evidence from Brazil."

Kang, Karam, and Bernardo S. Silveira. 2021. "Understanding Disparities in Punishment: Regulator Preferences and Expertise." *Journal of Political Economy* 129 (10): 2947–92. https://doi.org/10.1086/715415.

Lester, Benjamin, Ali Shourideh, Venky Venkateswaran, and Ariel Zetlin-Jones. 2019. "Screening and Adverse Selection in Frictional Markets." *Journal of Political Economy* 127 (1): 338–77. https://doi.org/10.1086/700730.

Luo, Yao, Isabelle Perrigne, and Quang Vuong. 2018. "Structural Analysis of Nonlinear Pricing." *Journal of Political Economy* 126 (6): 2523–68. https://doi.org/10.1086/699978.

Mahoney, Neale, and E. Glen Weyl. 2017. "Imperfect Competition in Selection Markets." *The Review of Economics and Statistics* 99 (4): 637–51.

Melo, Gilberto. 2015. "Lei Nº 13.043/14: Principais Alterações Nos Contratos de Alienação Fiduciária de Veículos." https://gilbertomelo.com.br/lei-no-13-043-14-principais-alteracoes-nos-contratos-de-alienacao-fiduciaria-de-veiculos/.

Mussa, Michael, and Sherwin Rosen. 1978. "Monopoly and Product Quality." *Journal of Economic Theory* 18 (2): 301–17. https://doi.org/10.1016/0022-0531(78)90085-6.

Pereira, Rafael H. M., and Daniel Herszenhut. 2025. *Geocodebr: Geolocalização de Endereços Brasileiros* (version v0.2.1).

Starc, Amanda. 2014. "Insurer Pricing and Consumer Welfare: Evidence from Medigap." *The RAND Journal of Economics* 45 (1): 198–220. https://doi.org/10.1111/1756-2171.12048.

Stole, Lars A. 2007. "Chapter 34 Price Discrimination and Competition." In *Handbook of Industrial Organization*, edited by M. Armstrong and R. Porter, 3:2221–99. Elsevier. https://doi.org/10.1016/S1573-448X(06)03034-2.

A Data construction

A.1 Sample selection

- Auto loans (code 0401)
- Only loans to households, residing in Brazil
- Funded by lender (excludes loans that were purchased from other lenders, or funded by earmarked funds)
- Fixed rate
- Only one collateral (the car)
- Following Joaquim, Doornik, and Ornelas (2023):
  - Loan start date later than borrower-lender start date
  - Loan start date earlier than loan end date
- Only cars: exclude motorcycles, trucks, buses
- Exclude loans with maturities with share inferior to 0.5%
  - Remaining maturities: 12, 18, 24, 30, 36, 42, 48, 60 months
  - 91% of the sample has maturities of 24, 36, 48, or 60 months

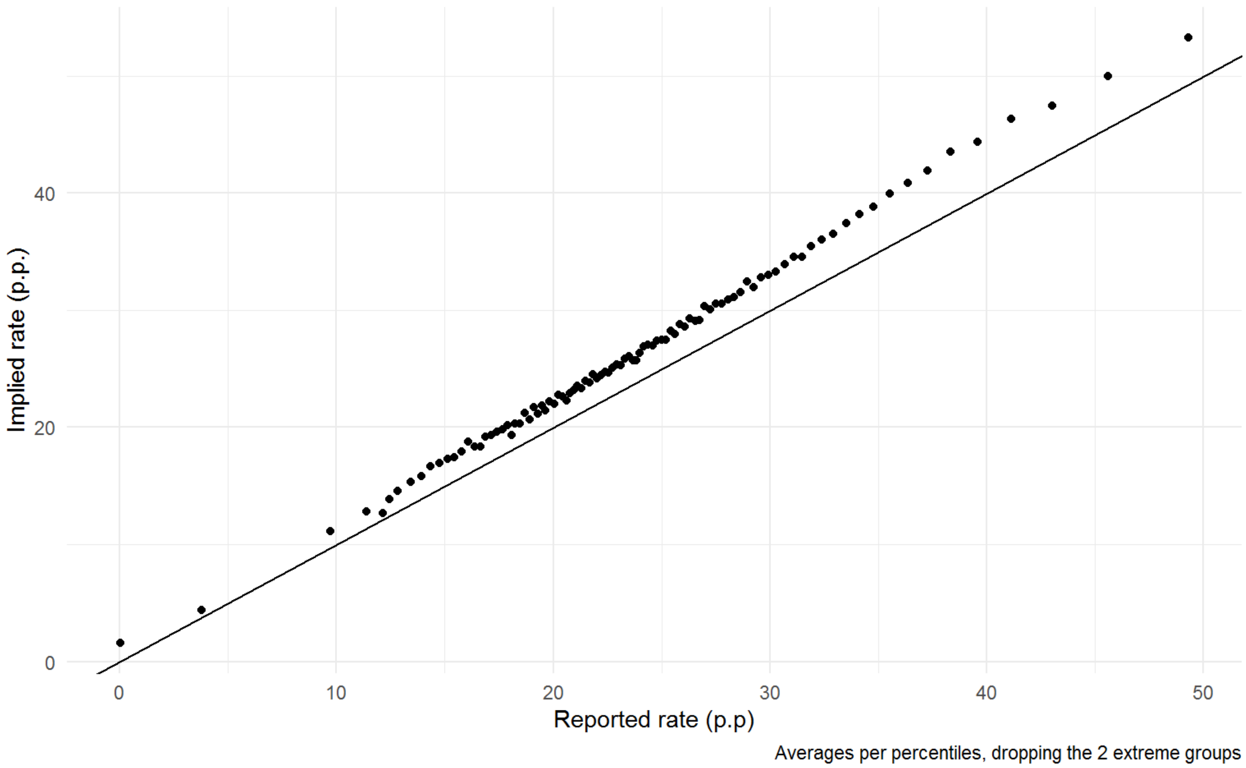
A.2 Interest rates

In this project, we are interested in what the Central Bank of Brazil calls the "effective rate", which includes all fees and taxes, and potentially other bundled products such as auto insurance. While lenders have to report interest rates for each loan, we found that there was heterogeneity in how they report it: in exploratory analysis, it seemed that some lenders would report the effective rate, while others would not.

To deal with this, we exploited that lenders also have to report the monthly payment, which is typically a fixed amount in the case of auto loans. Given that we also observe the loan maturity and the loan size, we can back out the loans' effective rate. We believe these values to be more reliable than the reported interest rates, as they must match the values in the lenders' balance sheets.

On Figure 3 we average the reported rate in percentiles, and plot on the y-axis the average implied rate for that percentile. We can see that implied rates are usually 2 p.p. higher than reported rates. This is consistent with some lenders not reporting the effective rate paid by borrowers. For this reason, we will adopt the implied rate as our main price measure, just discarding it when we detect issues with the reported monthly payments.

Figure 3: Comparison of interest rate variables



A.3 Income

We have access to two sources for borrowers' income: the income reported by lenders in the Credit Registry (SCR) and labor income from RAIS. While RAIS data is trustworthy, it has two main shortcomings: it accounts only for borrower-level income, not household level, which is the relevant measure for indebtness; it is available only for the population that has a formal salaried job, not including borrowers who are self-employed or work informally. On the other hand, the income reported by the lender, while available for almost the entire sample, is frequently misreported, as lenders' may not update frequently their customers' income. As an example, borrowers with an existing relationship with a lender had significanf lower reported incomes than those without a previous relationship, which may be a symptom of out-of-date information.

Therefore, we constructed a new dataset from the SCR containing the income reported by all lenders, not only the lender that issued the auto loan. Then, we defined borrower income as the maximum between the incomes available in the SCR and in RAIS. As a check, we can compare the income distribution in our sample to that from the Survey of Household Expenditures (POF 2016), as in Figure 4 and Figure 5. We can see that our measure is below the POF in the full sample, a result that is driven by the new car market. In the used car, we underestimate income for the left tail of the distribution but track closely at the upper quartile. In general, this comparison makes us take measures such as DTI and PTI with a grain of salt, as we are overestimating the weight of repayments in respect to income.

Figure 4: Comparison of income measures

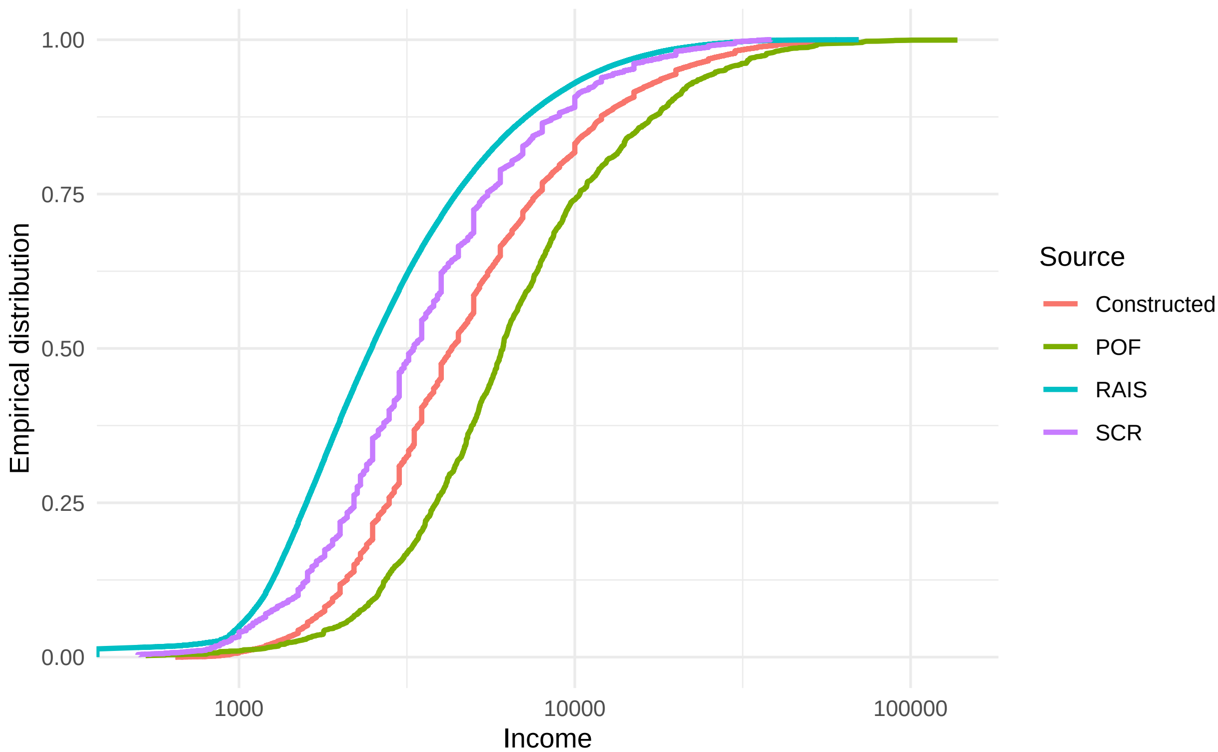
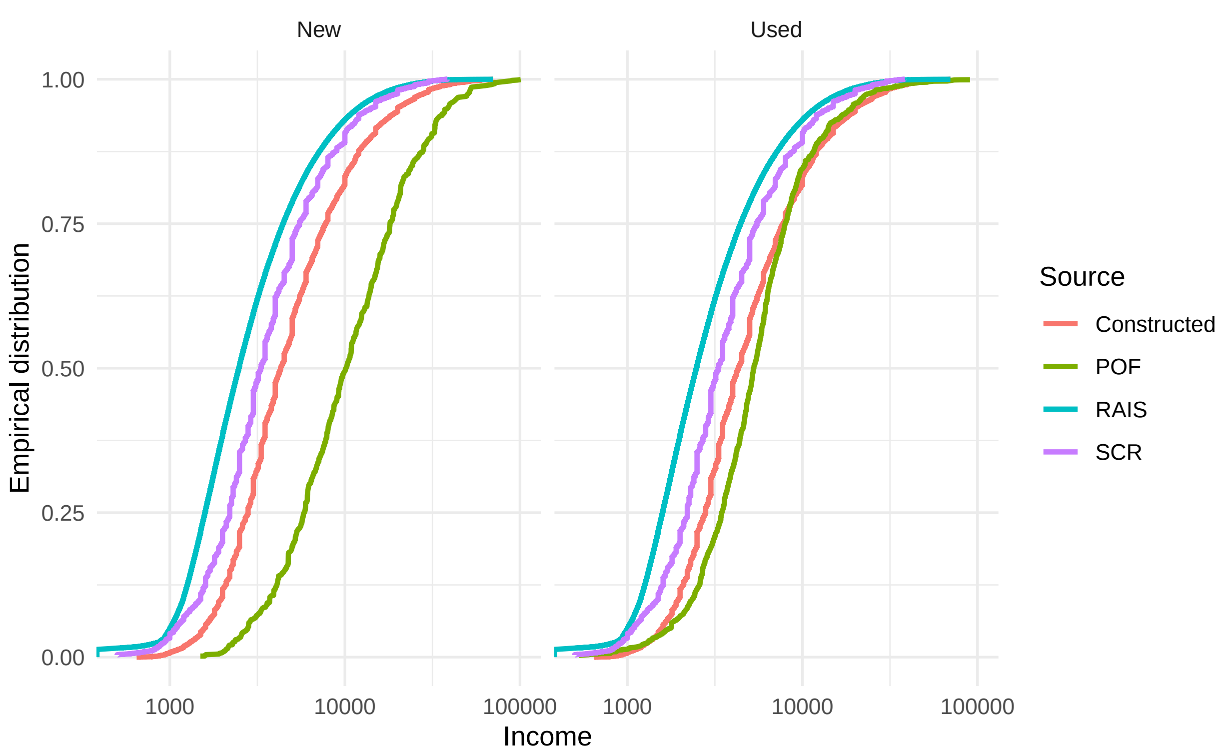


Figure 5: Comparison of income measures separately for new and used cars



A.4 Credit history

For each borrower, we reconstructed the information available to lenders through form 3046. We reconstructed the full information available for three months prior to the auto loan. Given that lenders have up to 24 months available, we took the average values for these months. The resulting dataset contains, for each credit type (e.g., credit cards, personal loans, credit limits, mortgages):

- Number of active loans
- Number of lenders
- Total amount outstanding
- Amount due in a given window (30/60/90/180/360/720/1080/1440/1800/5400/5400+ days)
- Amount in arrears in a given window (15/30/60/90/120/150/180/240/300/360/540/540+ days)
- Loan losses in a given window (12/48/48+ months)

Given the high dimensionality, we collapsed this information:

- Included one new category, summing across all loan types
- Kept separately only the most important loan types, by three criteria:
  - By total outstanding: car loans, mortgages, payroll loans, credit cards, and non-payroll personal loans (~90% of total outstanding)
  - By value due in 30 days: revolving accounts and revolving credit cards (~90% of total due in 30 days)
  - By losses: unpaid revolving credit cards
- Summary of credit limits: global, revolving accounts, and credit cards
- Collapsed the windows:
  - Amount due: within 30 days and within year
  - Arrears: within 30 days and over 30 days
  - Losses: total

In total, for each period there are 88 variables. Using 4 periods, three "standalone" and one 21-months average, we have 352 covariates. In Figure 6 we can see the most important loan types in our sample of auto loan borrowers.

Figure 6: Most relevant loan types by total amount outstanding and amound due within 30 days for the entire sample