Asymmetric Information and Search Frictions in Auto Loans*

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

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

This version: November 9, 2026
Click here for the latest version

Abstract

When borrowers have private information about their risk, safe borrowers can signal their type by borrowing less. Since the gain from a lower rate grows with the amount borrowed, borrowing less also weakens the incentive to search, softening competition among lenders. We measure the cost of asymmetric information, and of its interaction with search, in the Brazilian used-car loan market. Using the Central Bank of Brazil's Credit Registry, which records repayment and the lenders each borrower contacted, we find that larger loans are riskier conditional on lenders' information, and that borrowers who search more are riskier and pay higher rates. We estimate a model in which borrowers choose loan size and search effort given private search costs and a taste for borrowing correlated with risk. In the model, riskier borrowers are willing to accept higher rates, so they gain more from searching. We find that adverse selection has a small effect on average rates, as loan size reveals about half of borrowers' private risk. But signaling is costly: safe borrowers take smaller loans than under public information, and therefore search less. Making types public would raise total surplus by 22 percent, but by only 1 percent with loan sizes held fixed. Holding search fixed, the gain would be about a fifth smaller.

*We are grateful to Ken Hendricks and Jean-François Houde for their guidance and support, and to Jason Allen and Karam Kang for many helpful comments. We also thank our discussant Ron Young, an anonymous referee of the Central Bank of Brazil Working Paper Series, and participants at the UW–Madison IO Lunch, the 2026 Canadian Economics Association Meetings, the 2026 North American Summer Meeting of the Econometric Society, the 2026 EARIE Conference, and the 2026 Philadelphia Fed Conference on Auto Lending. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the Central Bank of Brazil. Contact: donedamittel@wisc.edu.

1 Introduction

In selection markets, consumers typically choose not only whether and from whom to buy, but also the contract characteristics, such as the deductible in insurance and the loan size in lending. This choice is informative: if riskier consumers self-select into lower deductibles and larger loans, firms can price the risk of each contract. In that case, lenders rationally charge higher rates on large loans, and safe borrowers can separate themselves by borrowing less (Milde and Riley 1988). Evidence from subprime auto lending is consistent with this form of selection (Adams, Einav, and Levin 2009). Signaling has a cost, however, because safe borrowers have to borrow less than they would otherwise. The signaling cost is larger the more steeply rates rise with loan size, which depends on how much lenders lose on risky loans. In our setting, the Brazilian used-car lending market, these losses are substantial, as 13 percent of loans default and lenders recover little of the collateral.

Search frictions add to the signaling cost. In markets with personalized pricing, lenders earn markups because obtaining competing quotes is costly for borrowers (Allen, Clark, and Houde 2019). Since the return to an additional quote is proportional to the amount borrowed, borrowers who reduce their loan size to signal low risk have weaker incentives to search, softening competition and increasing rates. As a result, signaling and search reinforce each other: part of what safe borrowers gain by signaling is lost to higher markups.

Our goal is to measure the cost of asymmetric information in this market, and how much the interaction with endogenous search matters. Existing approaches for consumer credit, building on Einav, Finkelstein, and Cullen (2010), take loan size and competition as given: borrowers only decide whether to borrow a given amount, and the degree of competition does not respond to how much they search (DeFusco, Tang, and Yannelis 2022). In our setting, both respond: safe borrowers borrow less to signal their type, and face weaker competition because they search less. If safe borrowers have already sorted into smaller loans, the remaining mispricing at a given loan size can be small even when the cost of asymmetric information is large.

We study this market using the Credit Registry of the Central Bank of Brazil, which records every used-car loan originated between March and December 2016, with repayment observed through 2024. Before 2021, Brazil had no widely used credit score, so the Registry was the main source of information lenders had about applicants. This allows us to reconstruct, for each loan, the information available to the lender at origination. Conditional on that information, borrowers who choose larger loans are more likely to default: a 10 percentage point increase in loan-to-value is associated with a 2.5 percentage point increase in default. This correlation could reflect either selection or a causal effect of debt

on repayment. To separate the two, we use variation in competition coming from changes in the lenders available at car dealers. Conditional on dealer identity, this variation is plausibly uncorrelated with borrowers' private information. When dealers add a lender, we show that borrowers pay lower rates and borrow more, but we find no evidence of changes in default rates. We interpret this result as evidence that the correlation between loan size and default is driven by adverse selection, not moral hazard.

The Registry also allows us to measure search. Each time a lender consults a borrower's record we observe it, so we know how many lenders a borrower contacted before taking the loan. Dealers play an important role: most loans are originated by the dealer on behalf of lenders in its network, and a borrower can obtain these quotes without searching on her own. Conditional on the information available to lenders, we find that borrowers who contact more lenders pay higher rates and default more often, and that this pattern is concentrated among observably risky borrowers. Among observably safe borrowers, more inquiries are associated instead with larger loans. Since search itself should lower the rate a borrower pays, we interpret these results as due to selection into search, with search incentives that vary with loan size and private information about risk. Separating these forces requires a model in which borrowers choose loan size and search together, and in which lenders price loan size as a signal of risk.

We develop a model of consumer loan-size choice, costly search, and lender competition. Borrowers differ in two dimensions that lenders do not observe: their valuation of borrowing, which is correlated with their default risk, and their cost of obtaining quotes. Each borrower chooses a loan size and a search effort. Lenders in her consideration set observe the loan size, form beliefs about her risk, and compete for the loan in an English auction, bidding rates that reflect their own costs (Allen, Clark, and Houde 2019). The model can generate the search facts above through the borrower's reservation rate. A borrower who values the loan highly will accept high rates. The higher her reservation rate, the more likely an additional quote is to fall below it and lower the rate she pays; for a borrower with a low reservation rate, search is more often wasted. Since a high valuation goes with high risk, riskier borrowers search more and still pay higher rates. As for loan size, the gain from a lower rate grows with the amount borrowed, so safe borrowers who borrow less to signal their type also have less to gain from searching.

We estimate supply and demand in two steps. First, on the supply side, we need to separate four determinants of the rate a borrower pays: the risk lenders expect, the number of lenders competing for the loan, their costs, and the borrower's reservation rate. We estimate lenders' beliefs about repayment directly from a model of default on loan size and the information in the Registry, which approximates what lenders knew at the time

of origination. The set of lenders competing in the auction is proxied by those who submitted an inquiry. To separate costs from reservation rates, we rely on exclusion restrictions that shift costs without changing the borrower's valuation, such as prior borrower–lender relationships. Given these, we can identify the cost distribution from the transaction rate and identity of the winner, following standard arguments (Athey and Haile 2002).

On the demand side, the main primitives are the distribution of borrowers' valuations, how valuation relates to risk, and the distribution of search costs. The supply estimates provide the rate distribution at each loan size. Search effort itself is not observed, and because borrowers choose loan size and search together, loan size and inquiries each reflect both their valuation and their search cost. We recover the distribution of types from how loan size, inquiries, default, and rates move together: a high valuation raises all four, while a low search cost raises inquiries but lowers the rate a borrower pays. The slope of default with respect to loan size identifies how closely valuation is tied to risk, and the dispersion of inquiries identifies the spread of search costs. Inquiries also mix the borrower's own search with quotes arranged by the dealer. We can separately identify each because we observe whether the loan was intermediated; if not, the borrower must have searched on her own.

The estimates point to a market in which both frictions are large. An additional quote costs the median borrower about R$746, or 3.5 percent of the median loan, consistent with borrowers contacting about two lenders on average. Lenders have considerable market power as a result: in most loans the rate is set by the borrower's reservation rate rather than by a competing offer. At our estimates, search frictions add almost 5 percentage points to the average rate. That is about half of the gap between what borrowers pay and what the most efficient lender would charge at cost; the rest is the markup lenders would earn even if all of them competed. Adverse selection, by contrast, has almost no effect on the average rate. This is not because private information is unimportant. Loan size reveals about half of borrowers' private risk. The remaining pooling only shifts rates between safer and riskier borrowers who take the same loan, by less than one percentage point a year (Appendix E.3). The cost of adverse selection therefore operates through quantities rather than prices.

Our counterfactuals quantify this cost. If lenders could observe each borrower's type, as they would with a sufficiently informative credit score, total surplus would rise by 22 percent, but by only 1 percent if each borrower's loan size were held at its baseline level. The amount lent would rise by about a third, mostly because borrowers who are now recognized as safe take larger loans. These borrowers also search more: holding search fixed, the gain in total surplus falls from 22 to 17 percent. Search frictions are also costly

on their own: removing search costs altogether would nearly double the market. Most of the cost of asymmetric information is therefore paid through smaller loans rather than higher rates.

A second counterfactual removes the signal itself. If lenders could not price loans by their size, borrowers would no longer need to signal through loan size, and they would borrow and search more. Consumer surplus would rise by 21 percent, and lender profits would also rise, by 14 percent, because larger loans more than offset the loss of information. Pricing by loan size is individually rational for each lender, but in equilibrium it lowers the surplus of both borrowers and lenders. This result depends on how lenders compete. In our model, lenders' bids vary with loan size only through expected repayment. Under different conducts, such as differentiated-product pricing, rates could also rise with loan size because borrowers who take larger loans are less price-sensitive, and restricting pricing by loan size would then also restrict this form of price discrimination (Nelson 2025).

Related literature. Our main contribution is to estimate a model in which loan size is both a choice and a signal of risk, and in which lenders' market power depends on how much borrowers search. Empirical models of credit markets with asymmetric information typically hold one of these margins fixed. Some take the loan amount as given and study the choice of lender or the decision to borrow (Cuesta and Sepúlveda 2021; Agarwal et al. 2024). Others let borrowers signal with the amount fixed, through the reserve rate on an online platform (Kawai, Onishi, and Uetake 2022). Einav, Jenkins, and Levin (2012) model the down payment as a screening device, but for a single lender, so competition plays no role. On the theory side, Lester et al. (2019) combine quantity screening with search frictions, but take the share of borrowers who search as exogenous. Letting both margins adjust matters for measurement. With a fixed loan amount, DeFusco, Tang, and Yannelis (2022) find that adverse selection distorts prices substantially but has a small welfare cost, and sufficient-statistic approaches measure that cost from the mispricing of a given set of contracts (Einav, Finkelstein, and Cullen 2010). The correlation between loan size and default used to test for asymmetric information (Chiappori and Salanié 2000; Adams, Einav, and Levin 2009) is not a measure of welfare either (Einav, Finkelstein, and Levin 2010). In our setting the pricing distortion is small for a different reason: borrowers have already sorted themselves by loan size. The welfare cost appears instead in how much safe borrowers borrow and, through that, in how hard they shop.

The idea that contract terms can separate borrowers goes back to the theory of credit rationing (Jaffee and Russell 1976; Stiglitz and Weiss 1981). Bester (1985) shows that when lenders set both the rate and collateral, borrowers self-select into separating contracts and

rationing disappears. Milde and Riley (1988) show that loan size itself can serve as the signal, as in our model, although in their case the direction of sorting is ambiguous. We find empirically that riskier types borrow more. The self-rationing of safe types resembles Guerrieri, Shimer, and Wright (2010), where good types accept lower trade probabilities to reveal themselves, and our equilibrium beliefs follow Azevedo and Gottlieb (2017) in requiring consistency at every loan size.

We also contribute to work on search frictions in credit markets. Our supply side follows Allen, Clark, and Houde (2019), in which lenders compete for each loan in an English auction. We add the choice of loan size, which makes the value of search depend on the amount borrowed. Agarwal et al. (2024) also use credit inquiries to measure search and find that riskier borrowers search more. In their model, this happens because risky borrowers are more likely to be rejected and search for approval. In ours, it comes from demand: riskier borrowers have higher reservation rates, so an extra quote is more likely to bring competition. Argyle, Nadauld, and Palmer (2023) show that high search costs lead auto borrowers to accept higher markups and take smaller loans. Our model adds the reverse direction, since borrowers who take smaller loans to signal have less reason to search. Dealers act as intermediaries in our setting, supplying quotes at no cost to the borrower. Salz (2022) shows that intermediaries can benefit both their users and, through a search externality, those who search on their own. Grunewald et al. (2023) show that dealers may also exploit what they learn about borrowers to add markups. We cannot observe dealer compensation, but we find that lenders in dealer networks have lower costs, which suggests that intermediation helps borrowers here.

Finally, the paper relates to work on how market power and adverse selection interact. Mahoney and Weyl (2017) show that with imperfect competition, more adverse selection has an ambiguous effect on prices, depending on the risk of the marginal consumer. Crawford, Pavanini, and Schivardi (2018) find that market power from lender differentiation attenuates adverse selection, because lenders internalize how their rates change the borrower pool. In their model the rate is the only screening device. In ours, borrowers signal through loan size, and market power comes from search frictions rather than differentiation. Nelson (2025) and Cuesta and Sepúlveda (2021) use models of this kind to evaluate price regulation, with borrowing as a binary choice. The interaction we study has a direct policy implication. Information policies such as credit scores and competition policies such as comparison platforms are usually evaluated separately. When loan size and search are linked, each also acts on the other friction: a credit score that lets safe borrowers stop signaling raises their return to search, and cheaper search lowers the rate on which the signaling premium is paid.

2 Institutional background

Credit markets in Brazil are characterized by high default rates, averaging 12.6% for used car loans in 2016. In contrast, in the same period in the United States, the average default was 3.8%.[1]

One friction that contributes to this market inefficiency is the absence of a widespread credit scoring system in Brazil before 2021. 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 whether 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 12,414/2011 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. Complementary 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 they reduced lenders' markups and increased the supply of credit (Banco Central do Brasil 2021; Hsu Rocha and Feinmann 2026).

A second friction in Brazilian markets is that, despite collateralized loans offering 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). Ratings for securities pooling used car loans also predict low recovery rates, ranging from 15% to 35% (Fitch Ratings 2014, 2025, 2026). While the existence of collateral still mitigates some of the risk for lenders, the low recovery rates still leave a big role for informational asymmetries to affect expected repayment.[2]

With respect to car sales, a customer has the option of buying a car with an upfront payment, financing, leasing, or through rotating savings and credit associations (*consórcios*, see van Doornik et al. 2024). For used vehicles, the share of financing over total sales was 21.8% in 2016 (B3 2020).

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

[2] Law 14.711/2023 allowed the repossession process not to require court approval, as in the United States, with the potential to streamline the repossession process and increase collateral value. For the analysis of a previous reform that improved repossession rates in Brazil, see Assunção, Benmelech, and Silva (2014).

3 Data

3.1 Data and summary statistics

Our main data source is the Credit Registry (SCR) of the Central Bank of Brazil, which stores the universe of loans issued by Brazilian lenders. Our sample includes all used-car loans to households originated from March to December 2016. We exclude from the sample loans for cars bought from individuals rather than firms. We observe loan 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. It 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. The structural model generates this menu; 3) we cannot directly identify customers who shopped for a car loan but did not get one. To account for this, we use total used-car sales as a proxy for market size.

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, which contains, for each month, a snapshot with the number of loans, number 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.[3]

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 with search intensity.

Inquiries should be interpreted as an imperfect measure of search, as some inquiries

[3]See Appendix A.3 for a description of this dataset.

may not be related to shopping for an auto loan. For example, many lenders submit inquiries monthly for all their clients to update risk models. We tackle this issue by constructing a measure of *previous inquiries*, defined in a window of more than 60 days before loan origination. We also observe that some lenders do not submit inquiries directly, instead routing them through partner institutions or using different data sources.

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.

In Table 1 we can see summary statistics of our dataset after all merges.[4] First, we see that the average interest rate, 29.7% per year, is much higher than the average interbank rate, of 13.0%.[5] We also observe that default rates, defined as a delinquency longer than 90 days, are high, averaging 12.6% over the loans' term. Intermediation, when auto dealers process the loan on behalf of lenders, is widespread in this market, accounting for 58% of all loans. The remaining loans are those where the borrower goes directly to a lender. Among intermediated loans, in 24.2% of cases the borrower had an existing bank account with the lender, and in 36.7% it had a previous loan. Among non-intermediated loans, these shares grow to 45.8% and 52.2%, showing that consumers value existing relationships.

From the inquiries dataset, we can observe that the median borrower has 2 unique lenders with inquiries in the 60 days before auto loan origination. Figure 1 shows the distribution: 90% of borrowers have inquiries from one to three lenders, and 2.5% from five or more. Given that we observe network status for each lender who submitted inquiries, we can measure that roughly half are submitted by lenders who are not in the dealers' network, and had to be reached by the borrower directly. As previously discussed, some of these can be related to previous loans. For this reason, we also calculate a "baseline" number of inquiries, measured between 60 and 90 days before loan origination, and therefore not related to the car acquisition.

For the estimation of the model, we need to define the total set of lenders available for each borrower and covariates that might influence their probability of being in the true consideration set or affect their costs. First, we define a market as a commuting region,

[4]For more details on sample selection and some variable definitions, see Appendix A.

[5]We use the future interbank rate evaluated at the relevant maturity for each loan. See more details on the construction of rate in Appendix A.2.

Table 1: Summary statistics of used-car loans (March–December 2016)

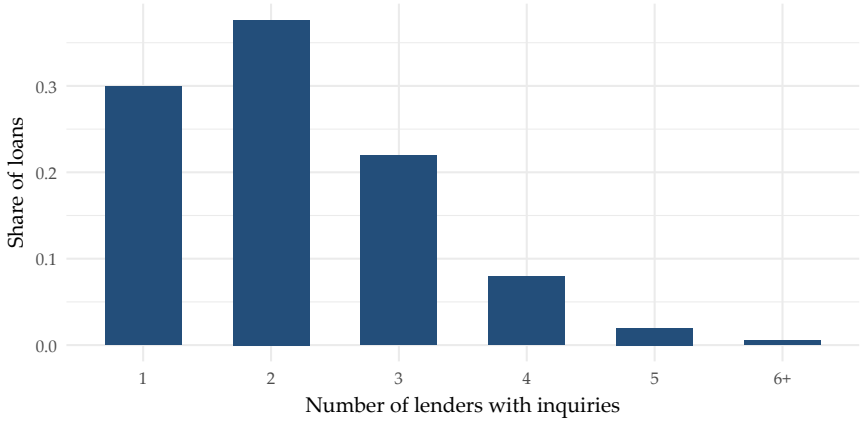
	Mean	SD	P25	P50	P75
Loan					
Interest rate (% APR)	29.69	7.52	24.31	28.15	33.32
Future interbank rate (% APR)	12.97	0.58	12.48	13.04	13.32
Loan amount (BRL)	23,666	12,056	16,047	21,200	28,000
LTV (%)	65.53	18.16	53.32	67.69	78.57
Loan term (months)	42.59	10.09	36.00	48.00	48.00
Default (1 = yes)	0.13	0.33	0.00	0.00	0.00
Intermediated (1 = yes)	0.58	0.49	0.00	1.00	1.00
Dealer network size	4.80	2.28	4.00	5.00	6.00
No dealer network (1 = yes)	0.09	0.29	0.00	0.00	0.00
Car					
Car value (BRL)	37,789	20,111	25,323	33,431	43,000
Car age (years)	4.57	2.73	3.00	4.00	6.00
Borrower					
Monthly income (BRL)	5,959	6,319	2,600	4,000	6,700
Predicted default risk	0.13	0.10	0.05	0.10	0.18
# inquiries (unique lenders)	2.17	1.05	1.00	2.00	3.00
# inquiries, in network	1.05	0.82	0.00	1.00	2.00
# inquiries, out of network	1.12	1.01	0.00	1.00	2.00
# lenders in the market	11.97				
Borrower–lender pairs					
In dealer's network (1 = yes)	0.37	0.48	0.00	0.00	1.00
Recent inquiry (1 = yes)	0.14	0.35	0.00	0.00	0.00
Banking relationship (1 = yes)	0.20	0.40	0.00	0.00	0.00
# previous loans	0.98	4.11	0.00	0.00	0.00
# previous car loans	0.10	0.38	0.00	0.00	0.00
Observations					
Loans	249,147				
Borrower–lender pairs	2,982,591				

Notes: Default: delinquency longer than 90 days over the life of the loan; within 12 months: the same event in the first 12 months after origination. Arrears: delinquency longer than 15 days. Inquiries: distinct lenders consulting the borrower's record in the 60 days before origination. Network size is zero when the seller could not be matched to an intermediary registered with the Central Bank.

and select only those that have more than 300 auto loans during the sample period, which results in 130 total markets.[6] For each market, we keep all lenders who issued more than 20 auto loans, and had a market share larger than 0.5%, covering more than 95% of total loans. On average, that gives 12 different lenders available per market. Among those, consumers had a previous banking relationship with 20%, a credit relationship with 17%, and 14% recorded an inquiry in the last 60 days before loan origination.

[6]There are 510 commuting regions (*região imediata*) defined by the Brazilian Institute of Geography and Statistics (IBGE).

Figure 1: Distribution of the number of inquiries



Notes: Lenders with an inquiry in the borrower's record in the 60 days before origination, top-coded at six. Shares from the data moments used in the demand estimation (Appendix D.4).

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 have private information over their risk, and riskier types borrow more. Adverse selection might coexist with moral hazard, defined in this context as the causal effect of loan size on repayment. The distinction is important, given that the policy recommendations for each are different, with credit scoring mitigating adverse selection and better repayment incentives being used to curb moral hazard (Adams, Einav, and Levin 2009).

Testing for the existence of asymmetric information is not straightforward in credit markets. Fixing contract characteristics, a test in the spirit of Chiappori and Salanié (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 our setting, given that we do not observe customers who choose not to borrow. Also, default is only defined for borrowers, such that we would need exogenous variation or parametric assumptions to compute counterfactual default rates for consumers who did not take a loan.[7]

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 larger loan-to-values (LTV) (Adams, Einav, and Levin 2009). Econometrically, we want to test if consumers

[7]For example, Crawford, Pavanini, and Schivardi (2018) assume a Heckman selection model in their extensive margin test.

Table 2: Boosted trees fit measures

Statistic	Baseline	Extended	Full
Default			
Brier score	0.101	0.097	0.096
ROC AUC	0.729	0.763	0.779
LTV (%)			
R-squared		0.516	
RMSE		12.640	
Interest rate (%)			
R-squared		0.729	0.762
RMSE		3.920	3.680

Notes: Out-of-fold fit: predictions are cross-fitted on five folds, with hyperparameters tuned by nested resampling. Baseline: pre-existing covariates only (income, sex, age, race, market, education, occupation, industry, job tenure, hours worked, credit history). Extended: baseline plus car value, brand, model, manufacturing and model year, loan term, lender and dealer identity. Full: extended plus the chosen LTV.

who choose higher LTVs (or, equivalently, smaller down payments) are more likely to default, after controlling for characteristics that affect pricing. Our basic specification for the asymmetric information test estimates two equations:

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

where X^L denotes loan characteristics, including number of installments, cohort, dealer ID; X^V includes vehicle characteristics, including car brand, model, trim, year, and value; X^C includes customer characteristics, including income, region, and demographics; X^{CH} includes the borrower's credit history[8], and X^B includes lender ID, conglomerate ID, and a dummy indicating a previous relationship between lender and borrowers. We adopt boosted trees, using the LightGBM algorithm, to approximate $f(\cdot)$ and $g(\cdot)$, and calculate $cor(\varepsilon, \eta)$, using the standard double/debiased ML approach (Chernozhukov et al. 2018). Table 2 contains measures of fit for each model, which were fitted using cross-validation on 5 folds, with nested resampling for hyperparameter tuning. The correlations were calculated considering the "extended" specification.

Correlation results are available in Table 3. The correlation is positive and statistically significant: an increase of a standard deviation in the residual LTV is associated with an increase of 0.100 standard deviation in the residual default risk. This correlation is 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. Alternatively, we can measure that an increase in LTV of 10 p.p. increases the probability of default by 2.46

[8]See Appendix A.3 for more details on the available variables

Table 3: Correlation test results

	N	Correlation	Default (pp) per 10 pp of LTV
All loans	249,147	0.100 (0.002)	2.46 (0.09)
Relationship			
Existing	110,731	0.090 (0.003)	2.03 (0.08)
New	138,416	0.108 (0.002)	2.83 (0.12)
Intermediation			
Intermediated	144,500	0.100 (0.002)	2.49 (0.13)
Not intermediated	104,647	0.100 (0.003)	2.41 (0.08)
Lender type			
Captive	23,049	0.096 (0.006)	1.77 (0.13)
Non-captive	226,098	0.101 (0.002)	2.56 (0.09)
Baseline risk quartile			
Q1	62,287	0.082 (0.004)	0.98 (0.04)
Q2	62,287	0.104 (0.004)	2.16 (0.09)
Q3	62,287	0.107 (0.004)	2.97 (0.13)
Q4	62,286	0.122 (0.004)	4.48 (0.15)

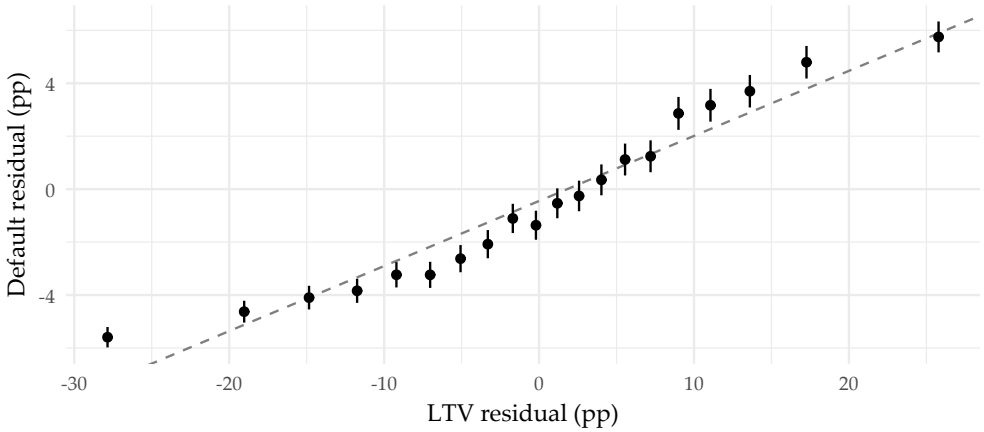
Notes: Correlation of the LTV and default residuals (extended specification), with influence-function standard errors. Slope: OLS of the default residual on the LTV residual, per 10 pp of LTV, standard errors clustered by region.

p.p. Considering an average default of 12.6%, this accounts for a 20% increase. In Figure 2 we can see that this effect is widespread over residual LTV. On Appendix B.1 we show that the association between loan size and default is robust to unobserved heterogeneity: controlling for the loan's interest rate, the results are similar.

The estimates are robust to excluding customers with an existing relationship with lenders, defined as having a loan or account with the lender at least one year before the auto loan: for customers that are new to the lender, the correlation is slightly higher, of 0.108. There is no difference in the correlation by intermediation status. We also do not observe a significant difference when comparing loans originated by captive lenders to non-captive lenders. Finally, as a sanity check for 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.

We note that the correlation results are consistent with different selection models. Our specification, which aligns with the structural model, assumes that consumers choose the LTV ratio, while holding car choice fixed. An alternative story is that consumers instead choose which car to buy. For example, borrowers have some amount of cash available, and riskier borrowers would self-select into more expensive cars in comparison to less risky borrowers, holding cash availability constant. This alternative model leads to the same estimation procedure, as the car choice is observable by the lender, and riskier borrowers would still be associated with higher LTVs.

Figure 2: Residual default and residual LTV



Notes: Twenty quantile bins of the LTV residual: bin means and 95% confidence intervals. Dashed line: linear slope of the default residual on the LTV residual.

One concern with the correlation test is that it does not allow us to disentangle adverse selection from moral hazard, as both produce the same positive correlation between loan size and default. To test which mechanism exists in the data, we need exogenous variation that affects loan size but is orthogonal to borrowers' private information. Our main instrument is the change in the number of lenders available at car dealers, which we call the network size. The assumption is that, conditional on dealer identity, changes in network are exogenous to consumers' private information on risk, and that it affects how much consumers borrow. Given that the effects scale with loan term, we add it interacted with the instrument.

In an alternative specification, we use average vehicle prices for the same model and year of manufacturing in other states. Its relevance is through how it affects the value of the car as a collateral for lenders, who can sell in the national market. The exclusion restriction is that the value of the car in other states does not affect how consumers repay the loan, conditional on market controls. We interact the instrument with borrower income (in log) to account for the fact that collateral value for the lender might differ depending on the borrowers' assets, proxied by income.

Results are in Table 4. The first column shows the correlation between default and LTV that we had already shown using the DML approach, here with OLS. In the other columns, we show the results using instrumental variables. Both instruments are relevant, as shown by the high F-statistic of the first stages (see Appendix B.2 for the first stage regression and the reduced-form estimation). The 2SLS estimates for the effect of LTV on default are slightly negative and not statistically significant at a 5% level. That is, we are unable to reject that the causal effect of loan size on default is zero. The confidence interval implies

Table 4: Adverse selection or moral hazard? Default on loan-to-value

		LTV instrumented		
	OLS	IV: network	IV: Hausman	IV: both
LTV (pp)	0.228***	−0.080	−0.089	−0.085*
	(0.009)	(0.055)	(0.082)	(0.048)
Observations	249,130	249,130	249,130	249,130
Mean default (share)	0.126	0.126	0.126	0.126
First-stage F		570.6	952.9	766.4
Sargan		0.086	0.087	0.181
Anderson–Rubin F		0.75	1.50	1.12

Notes: Dependent variable: default (pp). Network instruments: dealer network size and its interaction with log term. Hausman instruments: the average price of the same car in other states and its interaction with log income. Anderson–Rubin F: joint test that the instruments have no effect in the reduced form of default. The OLS column also controls for the interest rate. Controls: income, future interbank rate, DTI, car value, car spread, loan term, predicted risk, average monthly inquiries. Fixed effects: region, month, car age, model, brand, RAIS, occupation, seller, inquiry history. Standard errors clustered by region. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

that any positive effect, if one exists, is economically very small. We therefore find no evidence of moral hazard in this market, although we cannot rule out small effects: at the upper end of the 95% confidence interval, moral hazard would account for at most 12% of the OLS correlation in the specification with network size as instrument, and only 4% with both instruments. For this reason, our model assumes that adverse selection accounts for the entire residual correlation between loan size and default. These results are robust to adding the interest rate as a separate endogenous covariate. For more details, see Table B2 in Section B.2.

3.3 Search costs

Search costs are widespread in many markets, and have been extensively studied both theoretically and empirically. Search frictions are particularly relevant in a market with personalized pricing, given that consumers have to actively seek quotes from the lenders. In most settings it is not easy to measure such inefficiencies because search is not observed. We can leverage our dataset with inquiries to directly measure whether consumers with inquiries from more lenders obtain better terms.

One challenge is that not all inquiries are related to search for a car loan. For example, some inquiries might be related to previous loans, or to the lender's regular monitoring of their customers' credit history. Also, some inquiries might be related to search for other types of loans, such as mortgages or credit cards. To account for ongoing relationships, we calculate a "baseline" number of inquiries, measured between 60 and 90 days before loan origination, as well as the median number of inquiries in the 12 months prior to the

auto loan, and include both as fixed effects. As for simultaneous search for other loans, we treat it as random measurement error, so our estimates may suffer from the usual bias.

Denoting the number of total inquiries as N, we run the following regression:

$$y_i = \sum_{n \geq 2} \beta_n 1\{N_i = n\} + X_i' \gamma + \varepsilon_i$$

In outcomes y_i we include the interest rate, LTV, and default. In controls we include all the covariates described in the previous section, including loan characteristics, vehicle characteristics, consumer characteristics, and credit history, as well as month, dealership and baseline inquiries fixed effects. The coefficient of interest is β_n, which captures the correlation between search intensity and outcomes.

Results are available in Figure 3, where the baseline group are consumers with only one inquiry. The left-hand panel presents the results for the full sample and shows that search intensity is positively correlated with interest rates and default, similarly to the finding in Agarwal et al. (2024). We share their interpretation that this behavior is driven by selection of riskier types into search. While LTV is also increasing in search, we show in Appendix B.3 that the correlation with rates and default is not coming only from this channel, which we can show by controlling for LTV in the specifications with rate and default as outcomes.

Following the analysis in Agarwal et al. (2024), we run a separate set of regressions by splitting the sample into observably low-risk and high-risk consumers. Observable risk is estimated based on predetermined covariates, including demographics and credit history, from the baseline specification of the machine learning model from Section 3.2. The specification is

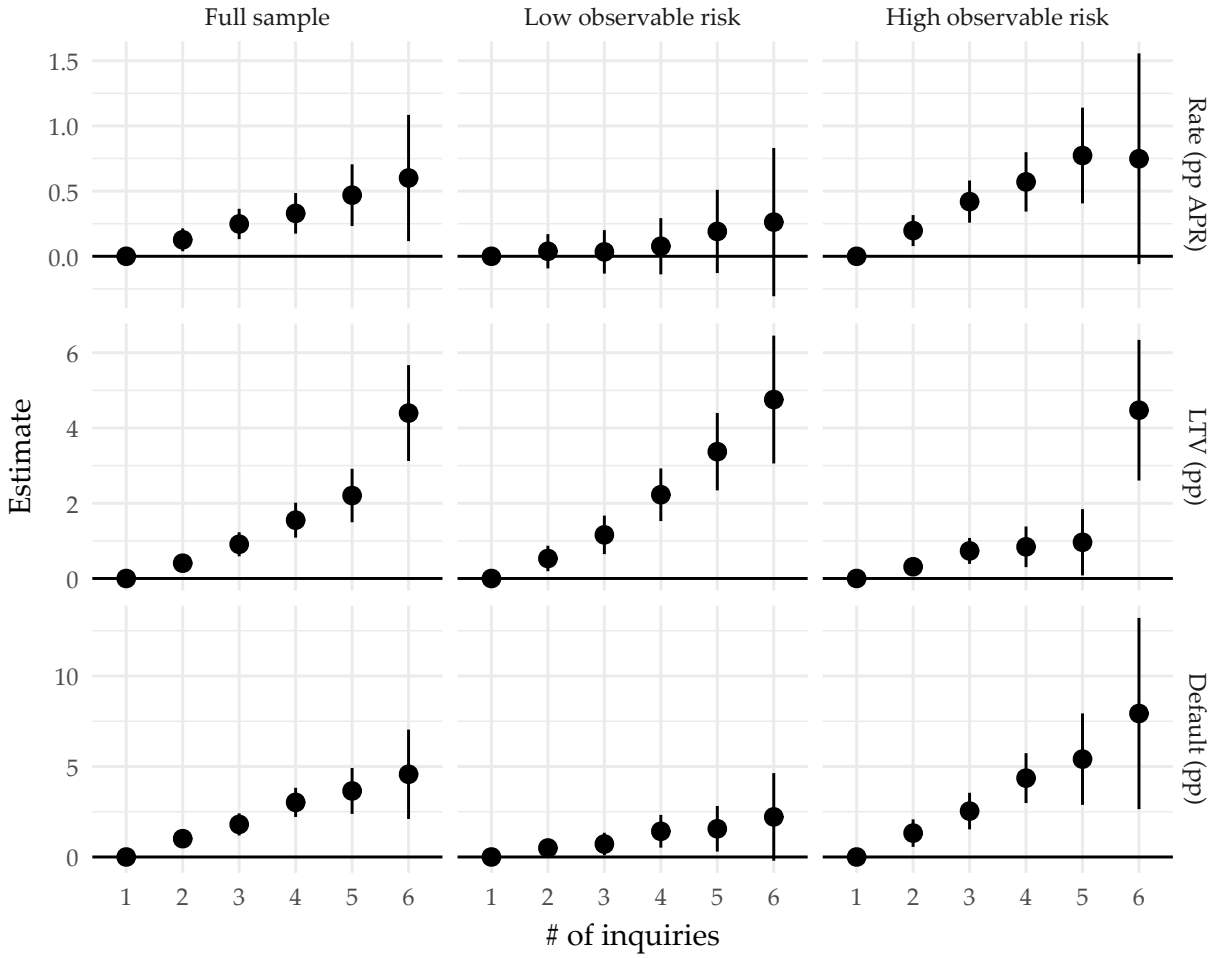
$$y_i = \sum_{n \geq 2} \beta_n^g 1\{N_i = n\} + X_i' \gamma^g + \varepsilon_i, \qquad i \in g,\ g \in \{\text{low risk}, \text{high risk}\}$$

The positive correlation is driven by observably high-risk consumers, and is essentially zero for low-risk consumers in terms of rate and much weaker for default. We take this as evidence that search incentives are increasing in risk, both observable and non-observable. While Agarwal et al. (2024) interpret this incentive as "searching for approval", we interpret it as coming from endogenous reservation rates.

4 Model

The model follows this timeline:

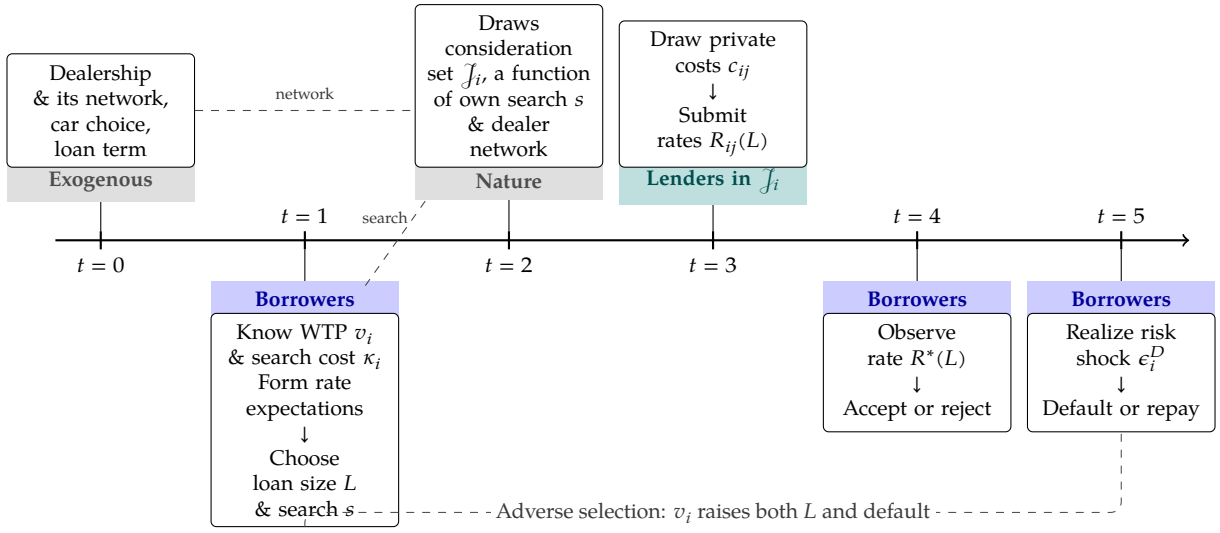
Figure 3: Interest rates, loan-to-value and default by number of inquiries (OLS)



Notes: Coefficients on inquiry dummies relative to one inquiry, with 95% confidence intervals; standard errors clustered by seller. Observable risk: above or below the median of predicted baseline risk, from boosted trees on pre-existing borrower characteristics only; separate regressions by group. Seven or more inquiries not shown.

0. Consumers exogenously choose:
    - Auto dealership, endowed with a network of lenders M_d for whom it can intermediate loans
    - Car, with fixed price V
    - Loan duration T, in months
1. Consumers draw a private taste for loan size v_i and their private search cost κ_i for each unit of effort, and, taking expectations over the interest rate for each option, they choose:
    - Continuous search effort $s \in \mathbb{R}^+$
    - Loan size $L \in [0, V]$

Figure 4: Model timeline



2. A Poisson process maps (s, M_d) into a consideration set $\mathcal{J}_i$
3. Lenders in the consumer's consideration set $\mathcal{J}_i$ draw costs c_{ij} and offer rates $R_{ij}(L)$, competing in an English auction
4. Consumers observe the outcome of the auction and choose whether to take the loan at rate $R^*(L)$ or the outside option U_0
5. Borrowers default or not

Figure 4 illustrates this timeline.

4.1 Default

Default is a binary outcome that reflects consumers' observed risk, their private information, and an idiosyncratic shock. We assume that consumers who do not default repay the loan in full, including principal and interest. Consumers who default might repay part of the loan. Formally, default is defined as

$$D_i = \mathbf{1}(\underbrace{X_i^D \beta^D}_{\text{Observed}} + \underbrace{\gamma^v \log v_i}_{\text{Private information}} + \underbrace{\epsilon_i^D}_{\text{Idiosyncratic}} > 0). \tag{2}$$

Adverse selection enters the model through a positive γ^v, which reflects the correlation of taste for loan size and risk. Covariates X^D are borrower characteristics that can affect default, including demographic information and credit history. They are observed by borrower and lenders. Idiosyncratic shock ϵ_i^D reflects uncertainty over repayment that is not accounted for by the previous terms.

Repayment expectations depend on the probability of default and the expectation over loss given default (LGD), which we denote by the function $\ell(X_i^D)$ and assume does not depend on borrowers' private information. For borrowers, who observe both covariates X^D and private information v_i, repayment uncertainty derives only from idiosyncratic shock ϵ_i^D. Lenders do not observe v_i, but are able to take into account selection effects by conditioning on loan size L and consumer acceptance of their quote. If $\gamma^v = 0$, borrowers and lenders share their repayment expectations. Mathematically, we denote repayment expectations for borrowers, $\pi^B(\cdot)$ and lenders, $\pi^L(\cdot)$ as

$$\begin{aligned} \pi_i^B(v_i, X^D) &= 1 - \ell(X_i^D)\Pr(D_i = 1 \mid X_i^D, v_i) \\ \pi_i^L(L, X^D) &= 1 - \ell(X_i^D)\Pr(D_i = 1 \mid X_i^D, L, \text{origination}). \end{aligned} \tag{3}$$

For conciseness, we use only π^B and π^L to denote the repayment expectations.

4.2 Loan acceptance

After observing the auction outcome, consumers can accept or reject the loan. They do so by comparing the utility they get from the loan, with amount L and rate $R(L)$, with the utility of the outside option U_0. Loan size was chosen in a previous stage, and is fixed here. The utility from a loan is given by

$$u_i(L, R) = \underbrace{v_i L}_{\text{Taste for loan size}} - \underbrace{\pi_i^B R L}_{\text{Expected repayment}} - \underbrace{\omega \Pr(D = 1 \mid v_i) L}_{\text{Default cost}} \tag{4}$$

It depends on the taste for loan size, which is consumer-specific through v_i. Consumers discount the utility by their repayment expectations, which depend on total charges, $R \cdot L$, and their assessment over how likely they are to repay it, π_i^B. Consumers also have a default cost, ω, which is common for all borrowers and accounts for the non-financial costs of defaulting, which includes the probability of losing the car and lower access to future credit. We calibrate ω at 0.03, three cents per BRL borrowed, rather than estimating it.

Optimally, consumers will have a threshold rule, such that they only accept loans with a rate lower than their reservation rate ρ_i:

$$\rho_i(L) = \frac{v_i - \omega \Pr(D = 1 \mid v_i) - \frac{U_0}{L}}{\pi_i^B} \tag{5}$$

Under $v_i > \omega$, we can show that $\frac{\partial \rho_i}{\partial v_i} > 0$: consumers with a higher taste for loan size

will also have higher reservation rates. Under adverse selection, consumers with high v_i are also riskier (more likely to default), such that riskier borrowers will have higher reservation rates.

The reservation rate in equation (5) allows us to rewrite the utility for a loan (L, R) as

$$u_i(L, R) - U_0 = \pi_i^B L \left(\rho_i(L) - R\right) \tag{6}$$

4.3 Rate determination (supply)

Lender j's expected profits for a loan to consumer i as a function of rates are given by

$$\mathbb{E}[\Pi_{ij}] = \underbrace{\pi_i^L(L) \times R \times L}_{\text{Expected revenue}} - \underbrace{(1 + c_{ij}) \times L}_{\text{Cost of funds}}, \qquad c_{ij} \sim F_j^C. \tag{7}$$

Revenue depends not only on principal and interest, but on the likelihood of repayment, denoted by π_i^L. The marginal cost of lending c_{ij} is a random draw from distribution F_j^C. We interpret this random cost as a match value, as in Allen, Clark, and Houde (2019) and Cuesta and Sepúlveda (2021). This match value can be due to the perceived value of establishing a relationship and the opportunities for cross-selling products.

We assume that lenders compete in rates as in an English auction, including all lenders in the borrower's consideration set $\mathcal{J}_i$. We assume that lenders observe consumers' loan size choice L, but not their search decision.[9] We also assume that lenders have the same information about consumers, such that their bids reflect only the idiosyncratic cost draw, ruling out a common-value component.

In an English auction equilibrium under private values, lenders bid the rate at which they break even:

$$R_{ij}(L) \;=\; \frac{1 + c_{ij}}{\pi_i^L(L)} \tag{8}$$

There are three potential outcomes: 1) when at least two lenders bid below the reservation rate, the runner-up's bid sets the transaction rate; 2) when only one does, the transaction

[9] Lenders in Brazil do not observe any information on consumer search behavior through the Central Bank's Credit Registry, in contrast with credit scores elsewhere, which flag when the consumer is shopping through the submission of "hard inquiries".

rate equals the reservation rate; 3) when none does, no loan is issued. Formally,

$$R^*(L) = \begin{cases} R_{\mathcal{J}(2)}, & \text{if } R_{\mathcal{J}(2)} \leq \rho_i \quad \text{(competitive)} \\ \rho_i, & \text{if } R_{\mathcal{J}(1)} \leq \rho_i < R_{\mathcal{J}(2)} \quad \text{(reservation rate binds)} \\ \emptyset, & \text{if } R_{\mathcal{J}(1)} > \rho_i \quad \text{(no trade)} \end{cases} \tag{9}$$

Under adverse selection, riskier borrowers prefer larger loans. Lenders take the selection into account in their repayment expectations π^L, which are then decreasing in loan size L. In equilibrium, rates take the form of a non-linear schedule depending on L.

4.4 Loan size and search

In this stage, consumers optimally choose a loan size $L \in [0, V]$ and a continuous search effort $s \geq 0$ to maximize their utility, taking expectations over the rate distribution. Using equation (6) and using the fact that consumers will optimally accept/reject loans in the later stage, we can rewrite the expected utility from a loan of size L and search effort s as

$$\max_{L \in [0,V],\, s \geq 0} \quad \pi_i^B L \; \mathbb{E}_{R^*}[(\rho_i(L) - R^*)^+] - \kappa_i s. \tag{10}$$

Consumers' utility from a loan depends on the difference between their reservation rate and the rate they get in the auction, scaled by loan size L and repayment expectations π^B. Consumers also incur a search cost κ_i for each unit of search effort. The loan size choice depends on the trade-off between the marginal gain from an additional unit borrowed and its cost, given a rate menu that is increasing in L.

In terms of search, consumers choose their search effort s optimally considering the search cost κ and the improvement in the rate distribution coming from search. They also take into account that they can have free quotes available at the car dealer, which depend on the number of lenders available. Assuming lender symmetry for simplicity, the map from search effort s into a number of quotes N follows a Poisson process

$$N \sim \text{Poisson}(\Lambda), \qquad \Lambda = \lambda M_d + s, \tag{11}$$

where parameter λ identifies the arrival rate of lenders in the dealer network, while the arrival rate of own search is normalized to 1. The benefit from search comes from increasing the number of lenders into the consideration set, which improves the expected rate distribution. On the other hand, consumers have to pay a search cost for increased effort. The free quotes provided at the dealers allow consumers to decrease their search effort.

Borrowers always have the option not to search, $s = 0$. If $M_d > 0$, they can still get a loan. If $M_d = 0$, they do not borrow and get the outside option, with utility U_0. Even if they search, there is a chance, given by the Poisson process, of getting no quotes, in which case they also get U_0.

In Appendix C.1, we present the first-order conditions for both L and s. In Appendix C.2, we cover the search technology and the first-order conditions when lenders are asymmetric.

The key mechanism in the model to reproduce the positive correlation between search and default in Figure 3 is given by equation (10). First, note that a quote that merely enables trade, $R = \rho_i(L)$, doesn't gain any utility over the outside option. In particular, given our supply-side conduct assumption, getting a single quote doesn't generate any consumer surplus. So there is no search incentive simply to get a loan, in contrast to Agarwal et al. (2024). Here, search incentives come from getting a quote that is below the borrower's reservation rate ρ_i. *Ceteris paribus*, consumers with a high reservation rate are the ones that gain the most from an additional quote. Given that ρ_i is increasing in v_i, and v_i is positively correlated with risk (adverse selection), riskier consumers benefit more from an extra quote. This is a mechanism through which the model can generate the pattern in Figure 3 for the full sample. The same reasoning works for observed heterogeneity in risk, as observably riskier consumers will also have a high ρ_i due to the lower probability of repayment π^B in equation (5).

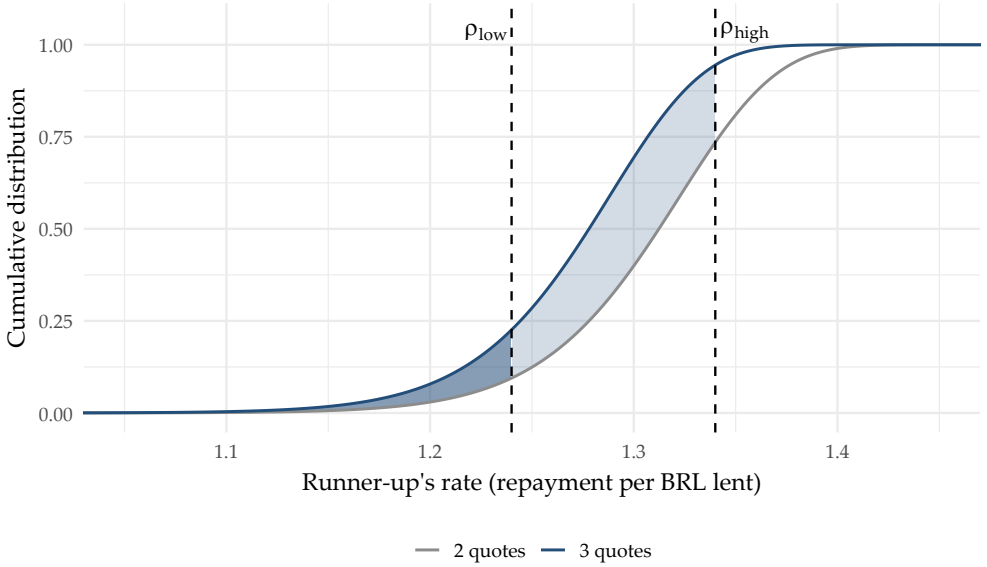
We can get more intuition about this mechanism in Figure 5, where we plot the transaction rate distributions for 2 and 3 symmetric lenders in the consideration set. High-risk consumers have a high reservation rate ρ_{high} and get a sizable benefit from adding a third quote, measured by the area between the two curves, whereas low-risk borrowers get a much smaller benefit, given their low reservation ρ_{low}.

4.5 Equilibrium

An equilibrium in this model is defined as:

- Consumers form expectations over rates and optimally choose s^* and L^*, given private info v_i and κ_i, car value V_i, network size M_d
- Lenders form expectations over repayment conditional on L and acceptance, given cost draws
- Lenders bid optimally in the English auction
- Consumers observe rates and optimally choose to accept/reject
- Consumers' rate expectations are consistent with how lenders set rates

Figure 5: The value of an additional quote increases with the reservation rate



Notes: Illustration with symmetric lenders at a given loan size. The gain from a third quote is the area between the two distributions below the reservation rate: dark when the reservation rate is low, dark and light when it is high.

- Lenders' repayment expectations are consistent with how consumers self-select into L

In Appendix C.3, we present lenders' repayment posteriors. We present the formal definition of the equilibrium in Appendix C.4.

We want to stress some features of the model. First, with a utility linear in loan size, if the rate menu were flat, all consumers would choose a corner solution, either not borrowing or borrowing the maximum amount available, a 100% LTV (Section C.1). In equilibrium, however, this doesn't happen, as lenders expect that high-risk borrowers will prefer larger loans, and their break-even bids form a nonlinear rate schedule increasing in loan size (8). It's this nonlinear menu that makes interior solutions available. The result is that low-risk consumers will distort their choice to signal their type, as in Bester (1985), while the highest-risk types will be unconstrained.[10]

Lenders' expectations at each loan size are consistent with the borrowers who choose it, as in Azevedo and Gottlieb (2017), but rates are set by an auction among the lenders in the consideration set, and therefore include a markup.

While in principle we could obtain a fully separating equilibrium, that will not necessarily happen. First, because there will be pooling at the top corner, among riskier borrow-

[10]This self-rationing is also similar to Guerrieri, Shimer, and Wright (2010), where good types accept worse trade probabilities to signal their type.

ers who choose the maximum LTV. Second, because borrowers differ also in search costs: consumers with low search cost will have access to more quotes and can borrow more than high-search-cost consumers. In our results in Section 6, we show that the redistributive effects of pooling are small, as the choice of loan size reveals much of borrowers' private information. The big cost of adverse selection comes instead from the signaling cost, as low-risk types have to borrow less than they would like under perfect information.

Importantly, the inefficiency caused by the signaling cost is amplified by search costs. Because search incentives grow with the amount borrowed, consumers who borrow less also search less, increasing market power and rates. In the counterfactuals, we compute how much market outcomes would change if types were observed, and decompose how much of the effect comes from loan size and search incentives.

5 Econometric model

The primitives of interest in the model are: 1) the distribution of consumer types, including the taste for loan size v_i and search costs κ_i; 2) the default parameters β^D and the adverse selection parameter γ^v; 3) the arrival rate of dealers' network lenders, λ; 4) the outside option U_0, in monetary value; 5) the non-financial cost of default, ω; 6) the cost distributions $F_j^C,\ \forall j \in \mathcal{J}$. In this section, we discuss the parametric assumptions we make to solve the model and how we allow for covariates.

5.1 Model specification

We assume that the taste for loan size v_i and the search costs κ_i are jointly lognormal:

$$\begin{pmatrix} \log v \\ \log \kappa \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} \mu_v \\ \mu_\kappa \end{pmatrix}, \begin{pmatrix} \sigma_v^2 & \rho_{v,\kappa}\sigma_v\sigma_\kappa \\ \rho_{v,\kappa}\sigma_v\sigma_\kappa & \sigma_\kappa^2 \end{pmatrix} \right) = \mathcal{N}(\mu, \Sigma). \tag{12}$$

In the estimation we set $\rho_{v,\kappa} = 0$, so search costs are independent of the taste for loan size. The idiosyncratic default shock is $\epsilon^D \sim N(0,1)$, with $\epsilon^D \perp (v, \kappa)$, so search costs are also unrelated to default risk. This multivariate normal specification allows us to have cleaner expressions for the lenders' beliefs, as in Adams, Einav, and Levin (2009). We do not let the parameters of the type distribution depend on covariates. Observed heterogeneity enters through the primitives of each borrower's problem: the car value, the loan term, the default index $X_i^{D\prime}\beta^D$, the lenders' costs and recovery, and the size of the dealer's network. We solve the model separately for groups of observably similar borrowers (Section 5).

On the supply side, we assume that the distribution of lenders' costs F_j^C is a Gumbel-min distribution, as in Allen, Clark, and Houde (2019) and Cuesta and Sepúlveda (2021), with location and scale parameters δ_C and σ_c. We allow the location parameter to depend on lender fixed effects, the future interbank rate, the loan term, the car's value and age, state fixed effects, and three measures of relationship: a previous bank account, the number of previous loans, and the number of previous auto loans. The scale parameter is constant across lenders.

5.2 Estimation and identification

For identification, we assume that vehicle, vehicle price, loan term, dealer, and dealers' network of lenders are exogenous. We assume that we have the same information set as the lenders (i.e., no unobserved heterogeneity), such that we can separately identify and estimate demand and supply. While the assumption might look strong, we rely on the fact that we reconstruct the information from the credit registry that was available to the lenders at the time of origination. We estimate the supply side first in a two-step procedure: first, we estimate repayment directly from the data, then we estimate the cost distribution using maximum likelihood. The rate distribution computed from these inputs is used to compute the gains from search: in expectation, by how much an additional quote reduces the expected repayment. This is the key ingredient we use to estimate demand, which we estimate with GMM targeting loan size, inquiries, rate, and default-related moments. The rest of this section describes the estimation procedure and how each parameter is identified.

5.2.1 Supply

Estimation of the supply side takes the distribution of borrowers as given, such that we can estimate expected repayment directly from the data.[11] The first step is to estimate lenders' repayment beliefs from equation (3).

We proceed in two parts: we first estimate a probit model of default on loan size and other covariates.[12] This probit model is similar to the more flexible machine learning

[11] Note that we do not need to assume at this stage whether a positive correlation between loan size and default comes from moral hazard or adverse selection, as from the lenders' point of view the relevant consequence is a lower probability of repayment.

[12] A fully consistent specification would also condition lenders' beliefs on the rate, as higher rates are accepted by a riskier pool of borrowers. We abstract from this effect, as it greatly complicates solving the model. Its magnitude is small: in the baseline equilibrium, lenders' expected profits, computed with beliefs that condition only on loan size, exceed the profits realized at borrowers' own default probabilities by 0.4%. While the gap is larger when we allow for more pooling, it reaches only 2.6% when lenders cannot condition

model in Section 3.2. Here we need extra structure, given that we use this model to extrapolate for each borrower the probability of default for different choices of loan size, ruling out a more flexible ML-estimation.[13] We then use the model estimates to predict for each borrower their probability of default given choice of loan size. Then, conditioning on default, we estimate the repayment as a fraction of principal and interest using a fractional logit model. We assume that only preexisting covariates, vehicle price, and loan term enter the LGD, not loan size. For details on the estimation of the probit and the fractional logit, see Appendix D.1. We then use the estimates to predict expected repayment given default for all consumers, under the assumption that, conditional on these covariates, repayment under default is independent of any private information held by consumers.

With these ingredients, we can rewrite the competitive supply equilibrium as

$$\widehat{\pi}_i^L(L)\ R_i^* \ = \ 1 + c_{(2)}^{\mathcal{J}_i}, \tag{13}$$

where $\widehat{\pi}_i^L(L) = 1 - \hat{\ell}(X^D)\ \widehat{Pr}(D|X^D, L)$ estimated from the probit and the fractional logit models for default and LGD, and R_i^* being the transaction rate. In this setting, with an observed set of bidders $\mathcal{J}_i$, we could identify the cost distributions F_j^C non-parametrically with the identity of the winner (Athey and Haile 2002; Komarova 2013). Instead, we follow the parametrization with a Gumbel-min distribution and estimate the supply with maximum likelihood.

Identification in our context is harder, given that the reservation rate can also determine the transaction rate when only one lender clears it, and it is unobserved by the econometrician. Instead, we rely on exclusion restrictions coming from our model: covariates such as a previous relationship with a lender do not enter the consumers' taste for loan size, but they affect the cost structure.

One key aspect of our data is the identity of the lenders who submitted an inquiry for each borrower, which allows us to proxy the actual consideration set. In practice, however, inquiries do not perfectly predict the bidders, as some may reflect shopping for other loans or existing relationships, as discussed in Section 3.3. We also identify some lenders that do not submit inquiries for loans they originated, such that an undercount is also possible. To account for this, we construct a measure of inquiry filing fraction conditional on winning per lender, and add this fraction to the likelihood, allowing it to adjust for a potential bias coming from conditioning on winning. See Appendix D.2 for more details and for the

on loan size.

[13] While we could use predictions from the machine learning model, such extrapolations are known to suffer from regularization bias. An alternative would be to use an orthogonalization approach, as in Chernozhukov et al. (2018). However, we are not aware of an orthogonal procedure for a probit model, given the need for probabilities bounded in $[0, 1]$, which rule out a partially linear model.

likelihood.

5.2.2 Demand

The demand model yields an optimal choice of loan size and search effort for each consumer type (v_i, κ_i). Integrating over the type distribution and over the Poisson search process, we obtain a distribution of loan sizes, quotes, and default that we compare with the data using GMM. In principle, we would need to calculate this distribution conditional on the observed heterogeneity, which includes observed risk covariates X^D and all covariates that affect demand parameters, such as income. To decrease the dimensionality of the state space, we use a k-means clustering algorithm to classify loans into K similar groups, and we solve the model separately for each group k, assuming homogeneity within group. For more details, see Appendix D.3.

In the demand model, we first assume that the asymmetric information is caused solely by adverse selection. That is, we assume no causal effect of loan size on default, consistent with our reduced-form finding in Section 3.2. The usual problem in this literature of decomposing asymmetric information between adverse selection and moral hazard is then innocuous, and the positive effect we observe of loan size on default directly drives the identification of parameter γ^v in equation (2).

The identification of the search costs comes from the dispersion in rates estimated in the supply side. More dispersion makes search more attractive, as each additional quote can significantly lower the transaction rate. Inquiry counts directly measure how much consumers are searching, and therefore how much they are "leaving on the table" by not searching in an extra lender, identifying the location of the search costs κ. Because the Poisson process ties the variance of the number of quotes to its mean, any dispersion in inquiries beyond what equal search effort would generate must come from heterogeneity in effort, and therefore in search costs: σ_κ is identified from the excess variance of inquiries relative to the Poisson benchmark.

The distinction between borrower and dealer-initiated search comes from the fraction of loans that were intermediated. If the arrival rate of in-network lenders were higher, consumers would take that into consideration and search less, increasing the fraction of intermediated loans. We also observe that in 9% of the loans in the sample the dealer had no network.

The mean and dispersion of the taste for loan size, μ_v and σ_v are identified from the loan size distribution. The utility of the outside option U_0 depends on the fraction of consumers who choose to borrow. Unfortunately, lender inquiries do not contain information on whether consumers were looking for an auto loan, such that we do not have microdata

on applications and rejections. Instead, we target the share of used-car sales that were financed.

For a full description of the moments and the weighting matrix, see Appendix D.4.

5.2.3 Model discussion

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 Sepúlveda (2021). The English auction assumption is mostly for tractability: a first-price auction, as in Salz (2022), would have similar outcomes, but it is much more complicated to solve its equilibrium under asymmetric lenders. Another alternative that relies on supply-side restrictions is given by D'Haultfœuille, Durrmeyer, and Février (2019), but their approach focuses on rate 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 rates, as in Crawford, Pavanini, and Schivardi (2018) and Brugués and De Simone (2024). However, there is significant residual dispersion in rates 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. (forthcoming), who design an outer fixed point that accounts for the missing rates and can be used within the differentiated products framework, but their approach was less tractable in our setting.

One important assumption is that lenders have symmetric information over borrowers' risk, and the information is common knowledge. This might not be the case if: 1) lenders have a private signal over each consumer's probability of default, in a common-values setting; 2) some lenders are differently informed about borrowers' risk. The second concern is easier to deal with, as long as the heterogeneity is common knowledge: we could simply add lender fixed effects in step 1 of the supply estimation, and solve the English auction equilibrium accounting for heterogeneous default beliefs, but we found the differences to be small. About the common-values setting, we believe that vehicle and loan size choices reveal enough of the borrowers' private information to make lenders' private signals second order. This would not be the case for uncollateralized loans in Brazil, for example, in which borrowers reveal much less information prior to origination.

A major difference from Agarwal et al. (2024) and theoretical papers on asymmetric

information is the absence of credit rationing and rejections. One difference is institutional: there are no rate ceilings in Brazil, such that even extremely risky consumers can still be profitable for lenders at very high rates. The absence of even informal ceilings is exemplified by rates that can exceed even 10% per month in uncollateralized loans. But, as in Stiglitz and Weiss (1981), even without rate ceilings there are settings where increases in rate can unravel the market. However, in our reduced-form estimates the direct effect of rate and loan size on default was too small to bend the revenue curve backward and generate endogenous rationing, and for this reason we shut down this channel for the estimation.

We also assume that search is not observed by the lenders, and therefore cannot be used to price the consumers' risk. We rely on the fact that the inquiry count was not available to the lenders through the credit registry. Nevertheless, we can imagine that out-of-network lenders could condition on the fact that consumers that reach out to them are unobservably riskier, as they have more incentives to search. This would create some heterogeneity in lenders' risk beliefs that, under common knowledge, could be added to our framework in future extensions.

More broadly, we can think that the borrowers' choices of car and dealer might also be indicative of their type, and our model takes those as exogenous. We believe that for the estimation this is the correct approach, as we should condition on the information set of the lenders. To the extent that adverse selection might be affecting also vehicle choice, such as with unobservably riskier borrowers choosing more expensive cars they do not expect to repay, we would be underestimating the degree of adverse selection. But given that this choice is observed by lenders, we do not expect it to affect our measurement of adverse selection conditional on observables, as we discussed for the correlation test. Vehicle choice might be relevant for the counterfactuals, though, as policy changes affecting access to credit might make consumers choose to buy a different car. We interpret our results with this concern in mind.

Regarding intermediation, there is a large literature showing that the benefits provided to consumers might not overcome misaligned incentives. In the case of auto loans, Grunewald et al. (2023) show that dealers might gather soft information on consumers to offer them more expensive credit, in exchange for kickbacks from lenders. Unfortunately, we do not have data on dealer compensation, and we did not find any direct evidence of such behavior in our context. For this reason, we do not directly add any costs of intermediation in our model. Nevertheless, our model allows for steering behavior, and we can compare the cost distribution of in-network and out-of-network lenders. As we show in our results, in-network lenders have lower costs on average, so dealers benefit consumers

in our setting.

Regarding search, one alternative framework would be a sequential search approach, as in Agarwal et al. (2024). We chose to model search as simultaneous mostly for tractability of the supply side. Sequential search would require solving a monopoly problem instead of an auction, breaking the identification of the cost distribution, as previously discussed. Nevertheless, we believe that both search models would bring the same incentive for consumers: in simultaneous search riskier types choose a higher search effort ex-ante, and get more quotes. In sequential search, riskier types would have lower acceptance thresholds and would search for longer before accepting an offer, generating the same predictions. In terms of functional form, we chose to add a Poisson process instead of a direct choice of the number of quotes. If lenders were symmetric, allowing borrowers to choose the number of quotes would be preferable, as in Salz (2022). However, if lenders are asymmetric their identity would matter, and the choice would be over the consideration set. This would greatly increase the choice set, which would grow to 2^{12} in the modal market, making it computationally infeasible. The Poisson-binomial process allows us to have random consideration sets in a tractable way, allowing lenders to be drawn at different rates depending on network status.

6 Results

6.1 Repayment estimates

The first step to estimate supply is to estimate repayment beliefs. The results from a probit model of default on LTV and other covariates are in the first column of Table 5. Consistent with Section 3.2, we observe a positive effect of LTV on default, with an average marginal effect of 0.3 p.p. on the probability of default for each additional percentage point of LTV. We proxy for the effects of the high-dimensional credit history with a risk index, calculated using boosted trees[14], which is mechanically highly predictive of default.

For repayment, we calculate for each defaulted loan the fraction of principal and interest that was repaid up to default, in present value. Then, we regress that on borrower and loan covariates, not including loan size. Finally, we project this prediction for every loan in the sample. Results are in the second column of Table 5. The prediction error is substantial, with a squared correlation of only 0.05, consistent with the notion that the month of default is very noisy. Essentially, the repayment expectation is mostly a function of the probability of default, not the amount repaid given default, and for this reason we focus on

[14]We use the *baseline* specification in Table 2.

Table 5: Lenders' repayment expectations – default and repayment given default

	Default	Share repaid | default
LTV (pp)	0.0175***	
	(0.0004)	
Predicted risk	3.4472***	−1.3259***
	(0.0622)	(0.0433)
log(Car value)	0.3248***	−0.1601***
	(0.0407)	(0.0139)
Car spread	0.5079***	
	(0.0367)	
log(Income)	0.0302***	−0.0107
	(0.0110)	(0.0086)
Future interbank rate	−0.0280	
	(0.0277)	
Debt-to-income (t-1)	−0.0001	
	(0.0001)	
Observations	236,200	41,619
Dep. var. mean	0.132	0.449
Pseudo R^2	0.193	
Squared correlation		0.046
Average marginal effect of LTV (pp)	0.304***	
	(0.006)	
Loan term fixed effects	Yes	Yes
Origination month fixed effects	Yes	Yes
Borrower, car, seller and lender fixed effects	Yes	No

Notes: Column 1: probit of default on the originated sample; coefficients on the index scale. Column 2: fractional logit (quasi-binomial likelihood, logit link) of the discounted share of the loan repaid, among loans that defaulted. Borrower, car, seller and lender fixed effects in column 1: region, RAIS, occupation, car age, model, brand, seller, lender. Standard errors clustered by region in column 1 (marginal effect by the delta method), IID in column 2. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

the binary outcome. Given the need to project the estimates also onto non-defaulters, we chose to use a parsimonious specification, without rich fixed effects. Results with richer specifications are not much different in terms of RMSE, and are available in Appendix D.1. One limitation of this approach is that we only use repayment up to the month of default, not accounting for payments after it, which might come from repossession of the vehicle or refinancing. For this reason, we calibrate the recovery given default using industry estimates, which range from 15% to 35% (Fitch Ratings 2014, 2025, 2026).

6.2 Cost estimates

Combining the default predictions from the probit model with the expected repayment given default, we can estimate the cost distribution with maximum likelihood using equation (13).[15] Results are in Table 6. We calculate that, on average, lenders' cost of lending

[15]For the estimation, we express rates and costs in present-value units per real lent: the rate R is the present value of the loan's scheduled installments per BRL of principal, discounted at the interbank rate, so

is 28% a year, with a dispersion of 3.5 p.p. This measured dispersion is close to the RMSE of the ML model we used to predict the interest rate.

We estimate costs excluding the top decile of predicted default risk. In that decile, observed rates respond much less to our default prediction than in the rest of the sample, even though the prediction is well calibrated there. This suggests that the prediction captures risk that lenders did not observe when pricing the loans, which the cost distribution would otherwise absorb.

The interbank rate does not have a meaningful impact on top of being the discount factor. Longer loans also do not affect costs. Car characteristics are relevant due to their value as collateral: older cars are costlier for lenders, and more expensive cars are associated with lower costs, as expected. Having an existing relationship with the borrower leads to lower costs, both for banking and credit relationships. These results proxy for cross-selling opportunities. We also observe large dispersion in costs among lenders, with a weighted IQR of 8.4 p.p. This is consistent with a market with heterogeneous lenders, as we observe both large banks with access to cheap deposits and smaller lenders with higher funding costs. Lenders in the dealers' networks are also cheaper: across borrower–lender pairs, the average in-network lender has a cost about 6 p.p. a year below the average out-of-network lender.

Market power is substantial: on average, the rate exceeds the winning lender's cost by 27% of the rate (Lerner index in Table 6). Despite that, markups are disciplined by the reservation rates: the estimates imply that in 63% of originated loans the rate is set by the borrower's reservation rate rather than by the runner-up's cost. The model also reproduces which lenders win: across lender groups, the RMSE of predicted and observed market shares is 3.5 p.p.

The full results including the probabilistic consideration sets and the distribution of reservation rates are in Appendix E.1.

6.3 Demand estimates

Using the cost estimates together with repayment beliefs, we construct a rate distribution and use it as input for estimating the demand, using GMM. The estimated demand parameters are in Table 7. These standard errors treat the first-stage estimates (default probit, repayment given default and supply) as known, and therefore understate the uncertainty in the demand parameters. The distribution of the taste for loan size shows that most consumers prefer to stay outside of the market, with a median willingness-to-pay of 14.4%,

that $R = 1$ means lending at the interbank rate. For presenting the results, we convert to APR.

Table 6: Lender cost estimates (annual percentage rate)

	Estimate
Mean cost (% APR)	28.00
SD of cost (pp)	3.55 (0.12)
Cost location (*pp of the annual cost rate*)	
Future interbank rate (pp)	−0.18 (0.07)
Loan term (+12 months)	0.17 (0.36)
Car age (years)	1.02 (0.05)
Car value (log)	−1.84 (0.36)
Banking relationship	−0.89 (0.09)
# previous loans (log(x + 1))	−1.19 (0.04)
# previous car loans (log(x + 1))	−2.19 (0.10)
Lender fixed effects (14 groups), weighted IQR	8.43
State fixed effects (25 groups), range	[−2.65, 0.47]
Lerner index (APR)	0.274
Win-share RMSE (lender groups)	0.035
Loans	212,649
Log-likelihood / N	−0.670

Notes: Maximum likelihood estimates of the Gumbel-min cost distribution, expressed as annual rates. Costs are estimated as the present value of repayments needed per BRL lent, discounted at the interbank rate, and converted to the annual rate of a 48-month loan at the average interbank rate (12.97%). Mean cost converted at the average loan term. Effects are changes in the annual cost rate; the interbank-rate effect is in addition to its role as discount rate. Loan term: change from a 48-month to a 60-month loan. Standard errors by the delta method, in parentheses. Lender fixed-effect IQR weighted by participation. Lerner index: average of (rate − cost) / rate over simulated originated loans, both as annual rates. Win-share RMSE: root mean squared difference between the observed and predicted shares of loans won by each lender group.

much smaller than the average lender cost. This is consistent with the low origination rate that we target, 33%. We achieve this estimate with the utility of the outside option being essentially zero.

In terms of adverse selection, we estimate $\gamma^v = 2.02$, implying a correlation of 0.26 between the taste for loan size and the default shock. We estimate a median search cost of 746 BRL per quote (216 USD), accounting for 3.5% of the median loan. In terms of dispersion, the 25th percentile borrower has a search cost of 715 BRL, and the 75th percentile of 779 BRL. For the median loan, the median search cost is equivalent to a reduction of 2.4 p.p. in the annual rate. Borrowers search until the expected reduction from one more quote falls to this value. Since each additional quote lowers the expected rate by less than the previous one, a search cost of this size is consistent with borrowers taking a small number of quotes. There are several potential explanations for the search costs. Under full rationality, they reflect the opportunity cost of borrowers for taking another quote, including the value of time and the hassle of starting a relationship with new lenders. Borrowers might also be unaware of the potential savings: for example, Grunewald et al. (2023) shows how some borrowers believe that the dealer always provides the best quotes.

Table 7: Demand estimates

	Estimate
Estimated	
WTP, median (% APR)	14.40 (0.05)
Dispersion of log WTP, σ_v	0.1351 (0.0007)
Corr(WTP, default), $\rho_{v,d}$	0.2634 (0.0045)
Implied adverse selection, γ^v	2.021 (0.044)
Search cost per quote, median (BRL)	746 (6.7)
Median, share of the median loan (%)	3.52
Median, rate equivalent (pp APR)	2.36
Dispersion of log search cost, σ_κ	0.064 (0.026)
Outside option, U_0, BRL	0.584 (0.962)
Dealer's arrival rate, λ	0.220
Calibrated	
Default cost per BRL, ω	0.030
Corr(search cost, default), $\rho_{\kappa,d}$	0
Corr(WTP, search cost), $\rho_{v,\kappa}$	0

Notes: GMM estimates; standard errors in parentheses, by the delta method for transformed parameters. Standard errors reflect sampling error in the moments only; the first-stage estimates (default probit, repayment given default, and supply) are treated as known. WTP (willingness to pay) is log-normal, in present-value units per BRL lent; its median is reported as an annual rate at the average loan term and interbank rate, in the same units as lender costs. $\gamma^v = \rho_{v,d}/(\sigma_v\sqrt{1-\rho_{v,d}^2})$. Search cost is log-normal in BRL. Its median is also reported as a share of the median loan (21,200 BRL) and as a rate equivalent: the reduction in the annual rate on the median loan, starting from the mean lender cost, whose present value equals the median search cost.

In terms of intermediation, we estimate that dealers do not use their entire networks: while an additional unit of search effort leads to one extra quote, one extra lender available creates only 0.2 extra quotes, on average.

The remaining parameters are calibrated: the correlations of search cost with taste for loan size and with risk are set to zero. We also assume a small non-financial cost of default, of 3 cents for each unit of loan.

6.4 Model fit

We assess the fit of the model in the first two columns of Table 8, which compare the data with the moments at the estimates, as computed in the GMM criterion. The model matches the correlation between loan size and default within groups of observably similar borrowers: the slope of default on loan size is 0.21 in the model and in the data, the moment that identifies the correlation between the taste for loan size and default risk. It also reproduces the number of lenders borrowers contact (2.23 against 2.15), with a default rate slightly above the observed one (14.5% against 13.2%). The average rate is 23.9% against 29.7%, as the rate moments receive a small weight in the estimation. The model does not reproduce the level of loan size in each group. The estimation targets the disper-

Table 8: Model fit

		Model	
	Data	At the estimates	Equilibrium
Default rate (%)	13.2	14.5	14.5
Mean LTV	0.655	0.630	0.723
Default–LTV slope, pooled	0.273	−0.082	−0.319
Default–LTV slope, within cluster	0.207	0.210	0.150
Number of lenders contacted	2.15	2.23	2.31
Rate (% APR)	29.7	23.9	27.0

Notes: Data: observed loans. At the estimates: the model moments at the demand estimates, as computed in the GMM criterion. Equilibrium: the counterfactual baseline, in which lenders' beliefs are solved to be consistent with borrowers' choices. Each column averages over its own distribution of loans within groups of observably similar borrowers (clusters), with groups weighted by their number of observed loans. Default–LTV slopes: least-squares slope of the default rate on LTV over LTV bins, pooling all clusters, or within cluster with cluster fixed effects. Number of lenders contacted top-coded at 6. Rates converted to annual rates at the average loan term.

sion of LTV within groups, but not how average LTV varies across groups, and the model places the riskiest groups at low LTV and the safest at high LTV (an average LTV of 0.46 against 0.69 in the data for the riskiest group, and 0.88 against 0.51 for the safest). As a result, and because default levels by group also differ from the data, pooling all groups, predicted default falls with loan size, while it rises in the data. Improving the fit of LTV in each group is an important next step for this work.

On the supply side, the model reproduces which lenders win: predicted market shares of lender groups differ from the observed ones by 3.5 p.p. on average (Table 6).

7 Counterfactuals

We use the estimates to simulate two sets of counterfactuals: 1) removing adverse selection; 2) removing search costs. Both are decomposition exercises that highlight the role of asymmetric information and search frictions in auto loans.

In the first set of counterfactuals, we study the effects of adverse selection by letting the private types v_i to be publicly observed by lenders, such that their beliefs over repayment are now the same as the borrowers', $\pi^L = \pi^B$. This counterfactual corresponds to the introduction of a credit score, for example, that would become a perfect signal of the borrower's private information about her default risk: lenders who observe it hold the same beliefs as the borrower, and price each loan at her own default probability. Formally, the counterfactual gives lenders a signal S_i that is a sufficient statistic for the borrower's private information about her default, $\Pr(D_i = 1 \mid X_i^D, S_i) = \Pr(D_i = 1 \mid X_i^D, v_i)$. Since private information enters default only through $\gamma^v \log v_i$ (2), the natural such signal is the

private component of the default index, $S_i = \gamma^v \log v_i$. A credit score that recovers this component, for example from the borrower's history with other lenders, would remove adverse selection entirely.

To understand the channels through which adverse selection operates, we also solve two variants of the first counterfactual, in which types are public but each borrower's search effort, or her choice of loan size, is held at the baseline while her other choices adjust. Finally, we consider a counterfactual in which lenders cannot condition their beliefs on loan size: they hold a single default belief for each group of observably similar borrowers, equal to the default rate among the group's originated loans, and use everything else they observe as before. This exercise corresponds to a rule that forbids pricing loans by their size, and isolates the role of loan size as a screening device.

In the second scenario we set $\kappa = 0$ for all borrowers, allowing them to search all available lenders in the market at no cost. While this is not directly a policy lever, search costs could be lowered by improving access to aggregators, for example, or favoring intermediation.

To compute the counterfactuals, we cannot hold lenders' beliefs fixed. Instead, we take the estimated parameters and solve the full equilibrium, requiring lenders' beliefs to be consistent with how borrowers self-select into loans, and therefore making the rate distribution also consistent with borrowers' rate beliefs. In practice, for each cluster used in demand estimation, we first solve the equilibrium under the baseline scenario, and then compare each of the counterfactuals to this baseline. Just as in demand estimation, we force consumers to choose LTV from a discrete grid. While this can be seen just as a computational simplification, we can also interpret it as proxying for other choice frictions, as in practice consumers bunch into LTVs that are multiples of ten. To facilitate convergence, we also add small logit shocks in the LTV choice, essentially making all LTVs on-path, and eliminating the need for further disciplining off-path beliefs, as in Azevedo and Gottlieb (2017). For computation, we use a combination of algorithms, including an Anderson-accelerated fixed point solver and a Newtonian solver.

Imposing consistent beliefs changes the baseline. In the estimation, lenders price loans with the repayment beliefs estimated from the data (Section 6.1); in the equilibrium, their beliefs must match the default of the borrowers who choose each loan size in the model. At the demand estimates, the two differ (third column of Table 8). The equilibrium shares the misfit across groups of the estimates: the riskiest groups remain at low LTV and default levels by group differ from the data, so that pooled default falls with loan size. Within groups, its gradient of default in loan size is about three quarters of the observed one (0.15 against 0.21), and the average rate is below the observed one (27.0% against 29.7%). The

misfit has two important consequences for the counterfactuals: the flatter gradient of loan size and default within groups should understate the adverse selection in the baseline. In addition, the misallocation of loan size across groups also changes how much each group's signaling cost weighs in the aggregate. Jointly, the effect on our counterfactual results is unclear. As discussed, improving both measures of fit is relevant for future work. The counterfactuals apply the cost estimates to all loans, including the top decile of predicted risk excluded from the supply estimation (Appendix A.1), which assumes that the cost function estimated on the other 90% of loans extends to that decile.

7.1 Results

Before analyzing the counterfactuals, our framework allows us to decompose the rate into different frictions. In a first-best scenario, the rate would be set by the lowest-cost lender, adjusted by the borrowers' own risk. To that, we add the market power coming from the auction, where the rate is set by the runner-up, and search friction, where it's set not by the runner-up across all lenders, but by the runner-up within the consideration set. Adverse selection also pools borrowers with different risks, adding an extra wedge. Ignoring the truncation by the reservation rate we can decompose the rates as follows:[16]

$$R_i^*(L) = \underbrace{\frac{1+c_{(1)}}{\pi_i^B}}_{\text{Mg. cost + risk adjustment}} \times \underbrace{\frac{1+c_{(2)}}{1+c_{(1)}}}_{\text{Auction market power}} \times \underbrace{\frac{1+c_{\mathcal{J}(2)}}{1+c_{(2)}}}_{\text{Search frictions}} \times \underbrace{\frac{\pi_i^B}{\pi_i^L(L)}}_{\text{Adverse selection}} \tag{14}$$

Table 10 reports the decomposition. Fixing search effort and loan size, in the first-best scenario the rate would fall to 17.5% a year, instead of 27.0%. The difference comes almost equally from auction market power (4.8 p.p.) and search frictions (4.7 p.p.). Adverse selection adds nothing to the average rate: since lenders' beliefs are consistent, pricing at a given loan size reflects the average risk of the borrowers who choose it, so borrowers safer than average pay slightly more and riskier ones slightly less. This cross-subsidy is small: it ranges from +0.8 pp a year for the safest decile to −0.5 pp for the riskiest, against a market-power component of 4.2–5.1 pp (Appendix E.3). Also, the equilibrium removes part of the pooling, as consumers with different risk choose different loan sizes, which is observed by lenders and priced in. We compute that loan size reveals about half of the variation in borrowers' private default risk among originated loans. In other words, adverse selection has strong implications in terms of who enters the market and how much they borrow, but conditional on borrowing a given amount, its effect is small, even taking

[16] See Appendix E.2 for the version accounting for the truncation.

into account the distributional effect.

Results for the adverse selection counterfactuals are in Table 9. With types being public, there would be an increase in total market size of 34%, from 6.6 billion BRL to 8.8 billion BRL. Part would come from an increase in the number of loans originated, of 13%, and part from larger loans, which would grow from 27.9 thousand to 33.1 thousand. Consumer surplus would increase by 13% and lenders' profits by 31%, raising total surplus by 22%.

Removing adverse selection also increases search: borrowers' own search effort rises by 12%. Once lenders price each borrower's own risk, rates no longer rise with loan size because of selection, and borrowers choose larger loans. As the search cost is fixed per quote while its benefit grows with the loan, larger loans make search more valuable. Private information therefore has an indirect cost beyond the pricing of risk: by discouraging large loans, it also weakens competition among lenders.

The interaction between adverse selection and search runs through loan size. Holding each borrower's choice of loan size at the baseline, making types public does not increase search: search per loan falls by 7%, as safer borrowers, who search little, now obtain lower rates and more of them take loans. Total surplus barely changes (+1%), so almost all the gains from removing adverse selection come from borrowers taking larger loans. Search amplifies these gains: holding search effort at the baseline, the increase in originations falls from 13% to 6%, and the increase in total surplus from 22% to 17%.

The last counterfactual in the adverse selection block shows that the inefficiency comes from the use of loan size to price risk. When lenders cannot condition their beliefs on loan size, but still use everything else they observe about the borrower, loans grow by 21%, as much as when types are public, and borrowers search 23% more. Originations fall slightly, by 2%, as safer borrowers now pay a rate that pools them with the riskier borrowers in their group. Total surplus increases by 17%, most of the gain from making types public, and consumers gain more than in that scenario (21% against 13%). Lenders also gain, by 14%: pricing by loan size is individually rational for each lender, given the information loan size carries, but in equilibrium it lowers the surplus of both borrowers and lenders. A more complete decomposition of total and per-loan surplus and profits is available in Section E.4.

The gains from removing search costs are even larger: the total market size would increase by 93%, divided between an increase in originations by 121% and a decrease in the average loan size. Given that search costs act as a fixed cost, removing them makes small loans more attractive to consumers. Total consumer surplus would triple, and lenders' profits would increase by 56%. Profit per loan falls by 30% as margins shrink, but this is more than offset by the larger number of loans.

Table 9: Counterfactual outcomes

						Surplus per loan (BRL)	
	Orig.	Market	L	Pr(D)	s^*	Consumer	Lender
Baseline	*1.00*	*6,583*	*27,869*	*0.145*	*1.29*	*1,947*	*2,246*
Adverse selection							
Public types	1.13	8,810	33,108	0.156	1.44	2,194	2,931
Search fixed	1.06	8,274	32,940	0.151	1.18	2,121	2,800
Loan size fixed	1.07	6,931	27,509	0.151	1.19	1,905	2,340
Pooling in loan size	0.98	7,826	33,646	0.146	1.59	2,358	2,558
Search frictions							
No search frictions	2.21	12,696	24,325	0.147	25.43	5,992	3,497

Notes: Orig.: originations relative to the baseline. Market: total amount lent, BRL million. L (loan size, BRL), Pr(D) and s^* (own search effort) are averages over originated loans. Surplus is per baseline originated loan; consumer surplus excludes the taste shocks in the choice of loan size. Lender: realized profit, at borrowers' own default probabilities. No adverse selection: lenders observe each borrower's type. Held at the baseline: each type keeps its baseline search effort, or its baseline choice of loan size, while the other choices adjust. Pooling in loan size: lenders hold one default belief per group of observably similar borrowers, the same at every loan size, equal to the realized default rate among the group's originated loans. All ten clusters, 236,200 loans.

Table 10: Rate decomposition in the counterfactuals (% APR)

		Decomposition		
	R^*	Cost + risk	Market power	Search
Baseline	*27.04*	*17.54*	*4.80*	*4.73*
Adverse selection				
Public types	27.27	18.18	4.84	4.25
Search fixed	27.15	17.83	4.82	4.51
Loan size fixed	27.14	17.85	4.82	4.48
Pooling in loan size	26.46	17.51	4.80	4.23
Search frictions				
No search frictions	22.47	17.68	4.76	0.05

Notes: R^*: transacted rate, averaged over originated loans. The decomposition splits R^* into the risk-adjusted cost of the most efficient lender, the markup from the auction among all lenders in the market, and the markup from the consideration set being smaller than the market, added in that order; the last two in pp of APR. The adverse-selection component is omitted: it redistributes among borrowers who take loans of the same size and averages to less than 0.1 pp. Rates are converted from present value per BRL lent to annual rates at the average loan term and interbank rate, as the lender costs. No adverse selection: lenders observe each borrower's type. Held at the baseline: each type keeps its baseline search effort, or its baseline choice of loan size, while the other choices adjust. Pooling in loan size: lenders hold one default belief per group of observably similar borrowers, the same at every loan size, equal to the realized default rate among the group's originated loans. All ten clusters, 236,200 loans.

8 Final remarks

This paper studies how adverse selection and search frictions jointly shape the market for auto loans in Brazil. Using administrative data from the Central Bank's Credit Registry, we provide direct evidence of both frictions and build a structural model that allows them to interact through endogenous loan size and search choices.

Our reduced-form evidence establishes two main findings. First, asymmetric information is pervasive: larger loans are at a much higher risk of default than smaller loans even once we control for rich borrower covariates. We show this effect is mostly driven by adverse selection rather than moral hazard. Second, we show that search is associated with higher rates and higher default, which is consistent with riskier borrowers having more search incentives. We take those findings into a structural model where borrowers choose loan size and search effort, with lenders competing in an auction.

Our model has features that are novel to the empirical literature on credit markets: low-risk consumers signal their type to lenders by self-rationing, mitigating adverse selection at the cost of lower borrowing. This cost is amplified by endogenous search, as smaller loans make search less attractive, increasing market power. Our empirical results show that the signaling cost is high, and a rate menu that is less steep in loan size can be welfare-improving, as consumers borrow more and search more.

These results highlight the importance of policies that aim to alleviate adverse selection by improving the information available to lenders, such as credit scores. We also show that search costs are high, and policies that facilitate price comparison and incentivize lender competition within auto dealerships can be welfare-improving.

Recently policies such as *Open Banking*/*Open Finance* aim to increase competition in credit markets by allowing borrowers to share their data with new lenders, decreasing informational asymmetries across lenders and reducing incumbency advantages. Our paper does not incorporate this dimension, as consumers signal their types through the loan size to all lenders, mitigating lender asymmetries in risk perception. In settings such as pre-approved personal loans, without a signaling component, lender differences in risk assessment should be more relevant. Incorporating lender informational asymmetries in a tractable framework remains an important avenue for future work.

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.

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.

Andrews, Isaiah, Matthew Gentzkow, and Jesse M. Shapiro. 2017. "Measuring the Sensitivity of Parameter Estimates to Estimation Moments." *Quarterly Journal of Economics* 132 (4): 1553–92.

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

Assunção, Juliano J., Efraim Benmelech, and Fernando S. S. Silva. 2014. "Repossession and the Democratization of Credit." *Review of Financial Studies* 27 (9): 2661–89.

Athey, Susan, and Philip A. Haile. 2002. "Identification of Standard Auction Models." *Econometrica* 70 (6): 2107–40.

Azevedo, Eduardo M., and Daniel Gottlieb. 2017. "Perfect Competition in Markets with Adverse Selection." *Econometrica* 85 (1): 67–105.

B3. 2020. "Mercado de financiamentos de veículos." Relatório mensal. B3 S.A. – Brasil, Bolsa, Balcão.

Banco Central do Brasil. 2021. "Análise dos efeitos do Cadastro Positivo." Report. Banco Central do Brasil.

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

Bester, Helmut. 1985. "Screening vs. Rationing in Credit Markets with Imperfect Information." *American Economic Review* 75 (4): 850–55.

Brugués, Felipe, and Rebecca De Simone. 2024. "Taxation When Markets Are Not Competitive: Evidence from a Loan Tax." STEG Working Paper 081. Centre for Economic Policy Research.

Campos, Álvaro. 2023. "Febraban mira veto à busca e apreensão de veículos." *Valor Econômico*, November.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. 2018. "Double/Debiased Machine Learning for Treatment and Structural Parameters." *Econometrics Journal* 21 (1): C1–68.

Chiappori, Pierre-André, and Bernard Salanié. 2000. "Testing for Asymmetric Informa-

tion in Insurance Markets." *Journal of Political Economy* 108 (1): 56–78.

Cosconati, Marco, Yi Xin, Fan Wu, and Yizhou Jin. Forthcoming. "Competing Under Information Heterogeneity: Evidence from Auto Insurance." *Review of Economic Studies*.

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 Sepúlveda. 2021. "Price Regulation in Credit Markets: A Trade-Off Between Consumer Protection and Credit Access." Working Paper 3999. Stanford Graduate School of Business.

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

DeFusco, Anthony A., Huan Tang, and Constantine Yannelis. 2022. "Measuring the Welfare Cost of Asymmetric Information in Consumer Credit Markets." *Journal of Financial Economics* 146 (3): 821–40.

Einav, Liran, Amy Finkelstein, and Mark R. Cullen. 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices*." *Quarterly Journal of Economics* 125 (3): 877–921.

Einav, Liran, Amy Finkelstein, and Jonathan Levin. 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics* 2: 311–36.

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

Fitch Ratings. 2014. "Fitch Atribui Ratings a 1ª Emissão de Cotas do FIDC Omni Veículos IX." March 26, 2014.

———. 2025. "Fitch Assigns Ratings to FIDC Creditas Auto XII Quotas." December 22, 2025.

———. 2026. "Fitch Assigns Ratings to the First Issuance of Quotas of FIDC Santander Auto Loans II." September 1, 2026.

Grunewald, Andreas, Jonathan Lanning, David Low, and Tobias Salz. 2023. "Auto Dealer Loan Intermediation: Consumer Behavior and Competitive Effects." NBER Working Paper 28136. National Bureau of Economic Research.

Guerrieri, Veronica, Robert Shimer, and Randall Wright. 2010. "Adverse Selection in Competitive Search Equilibrium." *Econometrica* 78 (6): 1823–62.

Hsu Rocha, Roberto, and Javier Feinmann. 2026. "The Effects of Revealing Borrowers' Information on Credit Allocation, Defaults and Entrepreneurship." Working Paper.

Jaffee, Dwight M., and Thomas Russell. 1976. "Imperfect Information, Uncertainty, and Credit Rationing." *Quarterly Journal of Economics* 90 (4): 651–66.

Joaquim, Gustavo, Bernardus van Doornik, and José Renato Haas Ornelas. 2023. "Bank Competition, Cost of Credit and Economic Activity: Evidence from Brazil." BIS Working Paper 1134. Bank for International Settlements.

Kawai, Kei, Ken Onishi, and Kosuke Uetake. 2022. "Signaling in Online Credit Markets." *Journal of Political Economy* 130 (6): 1585–1629.

Komarova, Tatiana. 2013. "A New Approach to Identifying Generalized Competing Risks Models with Application to Second-Price Auctions." *Quantitative Economics* 4 (2): 269–328.

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.

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

Milde, Hellmuth, and John G. Riley. 1988. "Signaling in Credit Markets." *Quarterly Journal of Economics* 103 (1): 101–29.

Nelson, Scott T. 2025. "Private Information and Price Regulation in the US Credit Card Market." *Econometrica* 93 (4): 1371–1410.

Salz, Tobias. 2022. "Intermediation and Competition in Search Markets: An Empirical Case Study." *Journal of Political Economy* 130 (2): 310–45.

Stiglitz, Joseph E., and Andrew Weiss. 1981. "Credit Rationing in Markets with Imperfect Information." *American Economic Review* 71 (3): 393–410.

van Doornik, Bernardus, Armando R. Gomes, David Schoenherr, and Jānis Skrastiņš. 2024. "Savings-and-Credit Contracts." SSRN Working Paper 4931994.

A Data construction

A.1 Sample selection

We start from all auto loans (SCR modality 401) to individuals. We keep loans funded with the lender's own resources and originated by it, which excludes loans purchased from other lenders and loans funded by earmarked resources, in Brazilian reais, at fixed rates, and with the car as the only collateral and no personal guarantee. Following Joaquim, van Doornik, and Ornelas (2023), we drop loans that start before the borrower's relationship with the lender or that end before they start. We keep loan terms with a share above 0.5% of loans (12, 18, 24, 30, 36, 42, 48 and 60 months) and loans with no missing key characteristics, such as the loan amount, term, income, and car model and value. Merging with the National Lien System (SNG) and the Federal Revenue, we keep cars, excluding motorcycles, trucks and buses, sold in a financed transaction, and borrowers residing in Brazil. We drop loans larger than twice the car's average market value, and trim the top and bottom 0.1% of the loan amount, the interest rate, income, debt-to-income ratios, car value and down payment.

Since the dealer-network data cannot be reliably merged before March 2016, we keep loans originated from March to December 2016, with at least one inquiry in the search window. A handful of borrowers take more than one car loan in the period; for them we keep only the first loan, or the largest if they took several on the same date. Table A1 reports the remaining steps. The largest drop comes from new cars. Supply estimation uses data on every borrower–lender pair in each market, defined as the borrower's region. We keep markets with more than 300 loans, which reduces the number of markets from about 500 to about 130, and, within each market, its main lenders: those with more than 20 loans and more than 0.5% of the market's loan volume. We also drop loans won by other lenders, and loans with missing values in the variables used in estimation, such as default. We also drop loans without a record of the borrower's inquiries before the search window, which we use to control for ongoing relationships. Finally, we drop loans with a debt-to-income ratio above the 97.5th percentile, and loans with an implied rate below the interbank rate.

The final sample has 249,147 loans, with an average default rate of 12.6%. It is used in the summary statistics and the reduced-form evidence. Regressions with fixed effects drop loans that are alone in their group. The default probit drops the fixed-effect groups in which no loan defaults, which carry no information on the effect of loan size, and is estimated on 236,200 loans, with an average default rate of 13.2%. Demand estimation and the counterfactuals use the same loans, since they rely on the probit's predictions. Supply

Table A1: Sample construction

Step	Loans	Dropped	Share kept (%)
Originated in 2016, sold by a car dealer	895,496		
Used car, with inquiries	320,657	574,839	35.8
Drop small markets	267,949	52,708	83.6
Inquiry history observed	265,038	2,911	98.9
Debt-to-income and rate filters	249,147	15,891	94.0

estimation excludes the top decile of predicted default risk (212,649 loans). Repayment given default is estimated on all defaulted used-car loans, except those for cars bought from individuals (41,619 loans), without the other restrictions above.

A.2 Variable definitions

Interest rate. The effective annual rate (*Custo efetivo total - CET*), which includes all fees and taxes, and potentially other bundled products such as auto insurance: the rate that equates the present value of the scheduled monthly payments to the loan amount. We compute it from the payment, the loan amount and the term rather than using the reported rate, as some lenders report rates before fees and taxes, while payments must match the values in the lenders' balance sheets. On average, the implied rate is about 2 p.p. higher than the reported one.

Car value and loan-to-value. Car value is the value of the car at origination, as reported by the lender for the loan's collateral, or, when it is missing, the car's reference price in the FIPE table, the standard table of average market prices by car model and year in Brazil, recorded in the SNG. LTV is the loan amount as a share of the car value. We do not observe the down payment directly, and impute it as the difference between the car value and the loan amount. The car spread is the ratio of the car value to its FIPE price.

Income. The borrower's monthly income, measured as the highest of the incomes reported by lenders in the Credit Registry and the formal wages in RAIS, as lender-reported income is often outdated and RAIS covers only formal salaried workers.

Default. A loan defaults if it is 90 or more days in arrears, or written off, at any point until the end of our performance data in 2024.

Inquiries. The number of distinct lenders, at the conglomerate level, that query the borrower's Credit Registry record in the 60 days before origination. An inquiry is in-network if the lender belongs to the dealer's network. Baseline inquiries are those 60 to 90 days before origination, and the inquiry history is the median number of monthly inquiries before that.

Intermediation and dealer networks. A loan is intermediated if a dealer originates it on the lender's behalf, as an intermediary registered with the Central Bank (*correspondente*). A dealer's network is the set of lenders for which the dealer is registered as an intermediary, and network size is the number of lenders in it.

A.3 Credit history

For each borrower, we reconstruct the information that lenders obtain when they query the Credit Registry before originating a loan (document 3046). Lenders observe up to 24 months of the borrower's history. We keep the three months before the auto loan separately, and average the remaining 21 months. For each month and credit type (e.g., credit cards, personal loans, credit limits, mortgages), the report contains the number of active loans, the number of lenders, the total amount outstanding, and the amounts due (within 30 days to more than 5,400 days), in arrears (from 15 to more than 540 days) and written off (within 12 months to more than 48 months).

Given the high dimensionality, we collapse this information. We add a total across all credit types, and keep separately only the most important ones: car loans, mortgages, payroll loans, credit cards, and non-payroll personal loans, which account for about 90% of the amount outstanding; revolving accounts and revolving credit cards, which account for about 90% of the amounts due within 30 days; and, by losses, unpaid revolving credit card balances. We also keep a summary of credit limits: overall, revolving accounts, and credit cards. Finally, we collapse the windows into amounts due within 30 days and within a year, arrears up to and over 30 days, and total losses. This leaves 88 variables for each period, and 352 covariates over the four periods.

B Reduced-form evidence

B.1 Correlation test controlling for the interest rate

The correlation test in Section 3.2 conditions on everything we observe, but lenders may observe more. If the correlation between loan size and default came from information that lenders have and we do not, lenders would price it, and controlling for the interest rate would absorb part of the effect. Using the residuals of the extended specification, we regress the default residual on the LTV residual, and then add the residual of the interest rate. Results are in Table B1. Each additional p.p. of LTV is associated with an increase of 0.25 p.p. in the probability of default, and adding the rate residual barely changes this

Table B1: Residual default on residual LTV, controlling for the residual interest rate

	(1)	(2)
LTV residual (pp)	0.246***	0.229***
	(0.008)	(0.007)
Interest rate residual (pp APR)		0.565***
		(0.024)
Observations	223,343	223,343
R^2	0.181	0.185

Notes: Dependent variable: default residual (pp). Residuals from the LightGBM extended specification. Controls: log income, DTI, future interbank rate, log car value, car spread, predicted risk, average monthly inquiries. Fixed effects: loan term, region, RAIS, occupation, car age, model, brand, month, seller, risk percentile, no network. Standard errors clustered by seller.

coefficient, to 0.23 p.p. The rate residual is itself associated with more default, which may reflect information that lenders have and we do not, or the selection of riskier borrowers into higher rates. Either way, it does not account for the correlation between loan size and default.

B.2 Adverse selection or moral hazard: additional results

This section complements the test for moral hazard in Table 4. Since larger loans also carry higher rates, which could affect default directly, in Table B2 we instrument both the interest rate and LTV, using the network and Hausman instruments together. Neither has a significant effect on default: the 2SLS estimates are -0.02 p.p. per p.p. of APR and -0.09 p.p. per p.p. of LTV, against 0.34 and 0.23 with OLS. The instruments are strong predictors of both, with first-stage F-statistics of 1,582 for the rate and 766 for LTV. The Sargan statistic, 0.13 with two overidentifying restrictions ($p = 0.94$), does not reject the validity of the instruments, and the Anderson–Rubin test ($F = 1.12$, $p = 0.35$) does not reject that both effects are zero.

The first stages are in Table B3, and the reduced form in Table B4. The instruments are not significant in the default regression, individually or jointly, consistent with the absence of a causal effect of loan size on default. We test an alternative specification with a control function probit instead of a linear probability model in Table B5, with similar results. Using the payment-to-income (PTI) ratio instead of loan size also does not change the conclusions.

Table B2: Default on the interest rate and loan-to-value, both instrumented

	OLS	2SLS
Interest rate (pp APR)	0.341***	−0.016
	(0.013)	(0.063)
LTV (pp)	0.228***	−0.088*
	(0.009)	(0.049)
Observations	249,130	249,130
Mean default (share)	0.126	0.126
First-stage F: rate / LTV		1,581.7 / 766.4
Sargan		0.125
Anderson–Rubin F		1.12

Notes: Dependent variable: default (pp). Column 2 instruments the interest rate and LTV with the network and Hausman instruments together. Controls, fixed effects and standard errors as in the main table. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table B3: First stages

	2SLS		LTV only		
	Rate	LTV	LTV: network	LTV: Hausman	LTV: both
log(Hausman inst.)	−7.743***	−8.316***		−8.401***	−8.316***
	(0.582)	(1.261)		(1.243)	(1.261)
log(Hausman inst.) × log(Income)	0.370***	3.150***		3.156***	3.150***
	(0.061)	(0.142)		(0.142)	(0.142)
Network size	−0.069	−4.739***	−4.772***		−4.739***
	(0.147)	(0.449)	(0.455)		(0.449)
Network size × log(# installments)	−0.188***	1.336***	1.341***		1.336***
	(0.037)	(0.119)	(0.120)		(0.119)
Observations	249,130	249,130	249,130	249,130	249,130
R^2	0.375	0.443	0.438	0.440	0.443
First-stage F	1,581.7	766.4	570.6	952.9	766.4

Notes: Dependent variable in the column header (rate in pp APR, LTV in pp). Controls and fixed effects as in the second stage; standard errors clustered by region.

Table B4: Reduced form of default on the instruments

	All four	Network	Hausman
log(Hausman inst.)	0.005		0.004
	(2.514)		(2.527)
log(Hausman inst.) × log(Income)	−0.226		−0.226
	(0.273)		(0.273)
Network size	0.411	0.414	
	(0.255)	(0.253)	
Network size × log(# installments)	−0.112	−0.112	
	(0.072)	(0.072)	
Observations	249,130	249,130	249,130
R^2	0.136	0.136	0.136
Instruments jointly zero (F)	1.1	0.7	1.5

Notes: Dependent variable: default (pp). Controls and fixed effects as in the second stage; standard errors clustered by region.

Table B5: Control-function estimates

	(1)	(2)	(3)
Interest rate (pp APR)	−0.016	−0.033	
	(0.058)	(0.058)	
LTV (pp)	−0.088**	−0.107***	
	(0.037)	(0.039)	
PTI		0.093***	0.091***
		(0.017)	(0.016)
First-stage residual, rate	0.363***	0.358***	0.326***
	(0.057)	(0.057)	(0.014)
First-stage residual, LTV	0.319***	0.311***	0.204***
	(0.040)	(0.039)	(0.008)
Observations	249,130	249,130	249,130
R^2	0.148	0.148	0.148
Dep. var. mean	0.126	0.126	0.126

Notes: Dependent variable: default (pp). Residuals from the two first stages (rate, LTV) on the four instruments. Controls and fixed effects as in the second stage; standard errors clustered by region, not corrected for the first stage.

B.3 Search regressions conditional on loan-to-value

Since LTV increases with the number of inquiries, part of the correlation between search and default in Figure 3 could come from searchers taking larger loans. Figure B1 repeats the regressions with the interest rate and default as outcomes, adding LTV as a control. The results are similar, and slightly larger for rates: in the full sample, borrowers with five inquiries pay rates 0.50 p.p. higher and are 3.2 p.p. more likely to default than those with a single inquiry, against 0.40 p.p. and 3.7 p.p. without the control. As in the main text, the correlation is driven by observably high-risk borrowers.

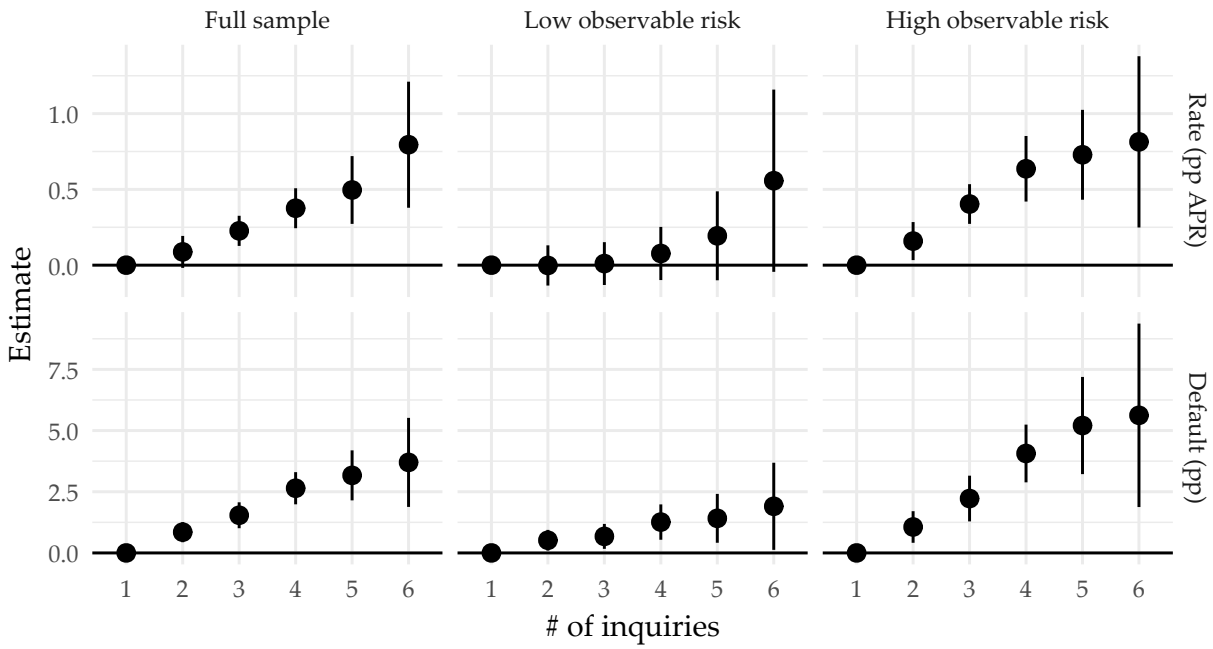
C Model derivations

Throughout this appendix we fix a borrower i and suppress the conditioning on observables X_i^D. The reservation rate equation (5) depends on the loan size through the outside option, so we write it as $\rho_i(L)$. Bids are $R_{ij} = (1 + c_{ij})/\pi^L(L)$, so lender j's bid is acceptable, $R_{ij} \leq \rho_i(L)$, if and only if its cost is below the reservation rate expressed in cost units,

$$z_i(L) \equiv \pi^L(L)\, \rho_i(L) - 1. \tag{C1}$$

We refer to $z_i(L)$ as the borrower's moneyness. Every object below depends on the type and on policies only through $z_i(L)$, the scale $1/\pi^L(L)$, and the arrival of quotes.

Figure B1: Interest rates and default by number of inquiries, controlling for loan-to-value (OLS)



Notes: Same specification as the main search figure, adding LTV as a control. Coefficients on inquiry dummies relative to one inquiry, with 95% confidence intervals; standard errors clustered by seller. Seven or more inquiries not shown.

C.1 Option value and first-order conditions with symmetric lenders

Option value. With symmetric lenders, costs are i.i.d. draws from F^C and the number of quotes is $N_i \sim \text{Poisson}(\Lambda_i)$, with $\Lambda_i = \lambda M_d + s_i$ (11). By equation (9), the borrower's payoff per unit of $\pi_i^B L$ is positive only when at least two bids are acceptable, in which case she pays the second-lowest bid:

$$(\rho_i(L) - R^*)^+ = \frac{1}{\pi^L(L)}(z_i(L) - c_{(2:N)})^+,$$

where $c_{(2:N)}$ is the second-lowest of the N_i cost draws, and the expression is zero when $N_i \leq 1$. For any random variable X with distribution F_X, $\mathbb{E}[(z - X)^+] = \int_{-\infty}^{z} F_X(t)\,dt$. By Poisson thinning, the number of quotes with cost below t is Poisson with mean $\Lambda_i F^C(t)$, so the probability that at least two costs are below t is $1-(1+\Lambda_i F^C(t))e^{-\Lambda_i F^C(t)}$. The expected option value is therefore

$$\Psi_i(L,s) \equiv \mathbb{E}\big[(\rho_i(L) - R^*)^+\big] = \frac{1}{\pi^L(L)} \int_{-\infty}^{z_i(L)} \Big[1 - (1 + \Lambda_i F^C(t))e^{-\Lambda_i F^C(t)}\Big]dt, \qquad \text{(C2)}$$

a one-dimensional integral, and the borrower solves $\max_{L,s} W_i(L,s) = \pi_i^B L\,\Psi_i(L,s) - \kappa_i s$.

Search. Since $\partial\Lambda_i/\partial s = 1$ and $\partial[1-(1+\Lambda F)e^{-\Lambda F}]/\partial\Lambda = \Lambda F^2 e^{-\Lambda F}$, the first-order condition for an interior search effort is

$$\frac{\pi_i^B L}{\pi^L(L)} \int_{-\infty}^{z_i(L)} F^C(t) \cdot \Lambda_i F^C(t)\, e^{-\Lambda_i F^C(t)}\, dt = \kappa_i, \tag{C3}$$

with $s_i = 0$ when the left-hand side evaluated at $s = 0$ is below κ_i. The integrand has a direct interpretation: $\Lambda_i F^C(t) e^{-\Lambda_i F^C(t)}$ is the probability that exactly one quote has cost below t, and $F^C(t)$ the probability that the marginal quote is also below t. An additional quote is valuable only when it is pivotal, converting a single acceptable bid, which leaves the borrower with no surplus, into competition between two. Because free quotes from the dealer raise Λ_i for any s, a larger network lowers the marginal value of own search once $\Lambda_i F^C(z_i)$ exceeds one. The left-hand side is proportional to L, so loan size and search are complements.

Loan size. Define the probability that the auction is competitive, $G_i^{(2)}(L,s) = 1 - (1+\Lambda_i F^C(z_i))e^{-\Lambda_i F^C(z_i)}$, and the expected rate paid in competitive outcomes, $\bar{R}_i^c(L,s) = \mathbb{E}[R_{\mathcal{J}^{(2)}} \mathbf{1}\{R_{\mathcal{J}^{(2)}} \le \rho_i\}]$, so that $\Psi_i = \rho_i(L) G_i^{(2)} - \bar{R}_i^c$. Loan size enters Ψ_i through the scale $1/\pi^L(L)$ and through the moneyness, with

$$\frac{\partial z_i}{\partial L} = \underbrace{\pi^{L\prime}(L)\,\rho_i(L)}_{\text{signaling}} + \underbrace{\pi^L(L)\,\rho_i'(L)}_{\text{outside option}}, \qquad \rho_i'(L) = \frac{U_0}{\pi_i^B L^2}.$$

Let $\varepsilon^L(L) \equiv -L\,\pi^{L\prime}(L)/\pi^L(L) \ge 0$ denote the elasticity of lenders' repayment expectations with respect to loan size, which is positive under adverse selection. Differentiating W_i and using $\partial\Psi_i/\partial z_i = G_i^{(2)}/\pi^L(L)$, the first-order condition for an interior loan size, $\partial \ln \Psi_i / \partial \ln L = -1$, can be written as

$$\varepsilon^L(L)\,\bar{R}_i^c(L,s) = \Psi_i(L,s) + G_i^{(2)}(L,s)\,\frac{U_0}{\pi_i^B L}. \tag{C4}$$

The right-hand side is the marginal benefit of borrowing an additional unit: the expected surplus per unit, and, when $U_0 > 0$, the relaxation of the acceptance constraint. The left-hand side is the marginal cost: a larger loan signals higher risk, which raises the rate on every unit borrowed in proportion to ε^L. Without adverse selection, $\varepsilon^L = 0$, and with $U_0 \ge 0$ the right-hand side is positive, so the borrower takes the largest loan available, $L = V$. An interior loan size requires that lenders price the risk revealed by the loan size. In the estimation, loan size is chosen on a discrete grid, and equation (C4) characterizes

the choice the grid approximates.

C.2 Asymmetric lenders

With asymmetric lenders, each lender $j \in \mathcal{J}_m$, with $M = |\mathcal{J}_m|$, draws its cost from its own distribution F_j^C, and the identity of the lenders in the consideration set matters. We model the formation of the consideration set through encounters. Borrower i meets lender j a number of times

$$E_{ij} \sim \text{Poisson}(r_{ij}(s_i)), \qquad r_{ij}(s) = \frac{s}{M} + \lambda \, \mathbf{1}\{j \in \mathcal{M}_d\}, \tag{C5}$$

independently across lenders, where $\mathcal{M}_d$ is the network of the borrower's dealer. Own search effort is spread over all lenders in the market, and the dealer generates additional encounters with the lenders in its network. Total encounters are $\sum_j E_{ij} \sim \text{Poisson}(\Lambda_i)$, with $\Lambda_i = \lambda M_d + s_i$, as in equation (11).

Each lender draws a single cost per borrower, so repeated encounters with the same lender do not generate new quotes. Lender j belongs to the consideration set if the borrower meets it at least once, which happens with probability

$$p_{ij}(s) = 1 - e^{-r_{ij}(s)},$$

independently across lenders. The number of quotes, $N_i = \sum_j \mathbf{1}\{j \in \mathcal{J}_i\}$, is therefore Poisson-binomial with success probabilities $\{p_{ij}(s_i)\}_j$, and $\mathbb{E}[N_i] = \sum_j p_{ij}(s_i) < \Lambda_i$. This is the object we measure with inquiries, which count distinct lenders. When lenders are symmetric and every p_{ij} is small, N_i is approximately $\text{Poisson}(\Lambda_i)$, which is the case in the main text.

Option value. Let $G_{ij}(t; s) = p_{ij}(s)\, F_j^C(t)$ be the probability that lender j is in the consideration set and has a cost below t. Membership and costs are independent across lenders, so the number of costs below t is a sum of independent Bernoulli variables with probabilities $G_{ij}(t; s)$, and

$$\Psi_i(L, s) = \frac{1}{\pi^L(L)} \int_{-\infty}^{z_i(L)} \Big[1 - \prod_j (1 - G_{ij}(t; s)) - \sum_j G_{ij}(t; s) \prod_{k \neq j} (1 - G_{ik}(t; s))\Big] dt. \tag{C6}$$

The integrand is the probability that at least two costs are below t. Its evaluation requires $O(M)$ operations at each t, with no sum over subsets of lenders.

Search. The integrand in equation (C6) is linear in each G_{ij}, and its derivative with respect to G_{ij} is the probability that exactly one other lender has a cost below t. Using

$\partial p_{ij}/\partial s = e^{-r_{ij}(s)}/M$, the first-order condition for search becomes

$$\kappa_i = \frac{\pi_i^B L}{\pi^L(L)} \sum_j \frac{e^{-r_{ij}(s)}}{M} \int_{-\infty}^{z_i(L)} F_j^C(t) \underbrace{\sum_{k \neq j} G_{ik}(t;s) \prod_{l \neq j,k} (1 - G_{il}(t;s))}_{\Pr(\text{exactly one other cost below } t)} dt. \tag{C7}$$

Search effort is valuable in proportion to how likely each lender is to be pivotal. Lenders with low costs are the most valuable to reach, and lenders that the borrower is likely to meet anyway, such as those in the dealer's network, contribute little at the margin, since $e^{-r_{ij}}$ is small for them. With identical lenders and small participation probabilities, equation (C7) reduces to equation (C3). The first-order condition for loan size keeps the form of equation (C4), with equation (C6) in place of equation (C2).

C.3 Lenders' posteriors

Lenders observe the requested loan size L, and are paid only by borrowers who accept their bid. Their repayment expectation in equation (3) conditions on both events. Let $\sigma_i(L)$ denote the probability that borrower i requests a loan of size L, and $\Pr(\text{orig} \mid i, L)$ the probability that she accepts a loan of that size. With symmetric lenders, the borrower accepts if at least one bid is below her reservation rate, so

$$\Pr(\text{orig} \mid i, L) = 1 - e^{-\Lambda_i^* F^C(z_i(L))},$$

where $\Lambda_i^* = \lambda M_d + s_i^*(L)$ evaluates her optimal search at L. With ϵ_i^D standard normal, the default probability of a borrower with private information v_i is $\Pr(D_i = 1 \mid v_i) = \Phi(X^{D\prime}\beta^D + \gamma^v \log v_i)$. By Bayes' rule, the default probability among borrowers who request L and originate is

$$\Pr(D = 1 \mid L, \text{orig}) = \frac{\int \Phi(X^{D\prime}\beta^D + \gamma^v \log v)\, \sigma_\theta(L)\, \Pr(\text{orig} \mid \theta, L)\, dF(\theta)}{\int \sigma_\theta(L)\, \Pr(\text{orig} \mid \theta, L)\, dF(\theta)}, \tag{C8}$$

where $\theta = (v, \kappa)$ is distributed according to F, and lenders' repayment expectation is $\pi^L(L) = 1 - \ell\, \Pr(D = 1 \mid L, \text{orig})$.

Selection operates through two margins. Riskier borrowers, with higher v, choose larger loans, which enters through $\sigma_\theta(L)$. They also have higher reservation rates, and therefore accept more often, which enters through $\Pr(\text{orig} \mid \theta, L)$. Both margins depend on $\pi^L(\cdot)$, through the rates the borrower expects and through the moneyness $z_i(L)$, so

equation (C8) defines a fixed point in the belief function $\pi^L(\cdot)$: in equilibrium, the beliefs that set the rates are the ones generated by the borrowers who choose each loan size under those rates. The posterior is undefined at loan sizes that no type chooses. In the estimation this does not arise, since lenders' expectations $\pi^L(L)$ are taken from the data rather than solved for (Section 5). In the counterfactuals, where beliefs must re-equilibrate, we add logit shocks with scale λ^L to the loan-size choice, as in Azevedo and Gottlieb (2017), so that

$$\sigma_\theta(L_k) = \frac{\exp \frac{W_\theta(L_k)}{\lambda^L}}{\sum_m \exp \frac{W_\theta(L_m)}{\lambda^L}},$$

with $W_\theta(L) = \max_s W_\theta(L, s)$. Every loan size on the grid is then chosen with positive probability, and beliefs are defined at every grid point.

Because loan size is chosen from a finite grid of Q points with logit shocks, the map from lenders' default beliefs at each grid point, a vector in the compact convex set $[0,1]^Q$, to the default rates those beliefs induce among borrowers who choose each L and accept is a continuous self-map. Brouwer's theorem therefore guarantees a fixed point.

C.4 Equilibrium definition

Given the parameters and the observed characteristics of a cluster, an equilibrium consists of borrowers' search policies $s^*_\theta(L)$, loan-size choice probabilities $\sigma_\theta(L)$, and lenders' repayment expectations $\pi^L(L)$, such that: (i) given $\pi^L(\cdot)$, lenders bid their break-even rates (8), and rates are determined by equation (9); (ii) given the resulting rate distribution, each type θ chooses search effort at each loan size to maximize equation (10), chooses loan size with the logit probabilities in Appendix C.3, and accepts the loan if the rate is below her reservation rate; and (iii) given these choices, $\pi^L(L)$ equals the repayment rate among borrowers who choose L and originate (C8), at every loan size on the grid.

D Estimation

D.1 Supply-side inputs

Table D1 adds the interest rate to the default probit in Table 5. Higher rates are associated with more default, but including the rate barely changes the effect of LTV, with an average marginal effect of 0.29 p.p. per p.p. of LTV, against 0.30 p.p. in the baseline. Table D2 adds a rich set of fixed effects to the repayment model. The squared correlation increases from 0.05 to 0.14, but the RMSE decreases only from 0.27 to 0.26, and the coefficients are similar.

Table D1: Probit of default – baseline and with the interest rate

	(1) Baseline	(2) With rate
LTV (pp)	0.0175***	0.0167***
	(0.0004)	(0.0003)
Interest rate (pp APR)		0.0304***
		(0.0008)
Predicted risk	3.4472***	3.4114***
	(0.0622)	(0.0647)
log(Car value)	0.3248***	0.4346***
	(0.0407)	(0.0398)
Car spread	0.5079***	0.4554***
	(0.0367)	(0.0363)
log(Income)	0.0302***	0.0281***
	(0.0110)	(0.0106)
Future interbank rate	−0.0280	−0.0336
	(0.0277)	(0.0272)
Debt-to-income (t-1)	−0.0001	−0.0000
	(0.0001)	(0.0001)
Observations	236,200	236,200
Dep. var. mean	0.132	0.132
Pseudo R^2	0.193	0.200
Average marginal effect of LTV (pp)	0.304***	0.288***
	(0.006)	(0.006)
Average marginal effect of the rate (pp)		0.523***
		(0.014)

Notes: Dependent variable: default. Probit coefficients on the index scale. Column 1 is the specification in the main text. Both columns include loan-term dummies and fixed effects for region, origination month, car age, model and brand, RAIS, occupation, seller and lender. Standard errors clustered by region; marginal effects by the delta method. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Finally, Figure D1 compares predicted and realized default by LTV. The probit reproduces the increase in default with LTV, including its flattening above 80% LTV. For a loan with average characteristics, the probability of default increases from 4.8% at 50% LTV to 12.7% at 80% and 21.5% at 100%.

D.2 Supply likelihood

We estimate the cost distributions by maximum likelihood, using the transacted rate, the identity of the winner, and the lenders with inquiries. Following equation (9), the rate of a loan is set either by the runner-up's bid (competitive branch) or by the borrower's reservation rate (reservation branch), which we do not observe. Conditional on origination, the likelihood of loan i is below.

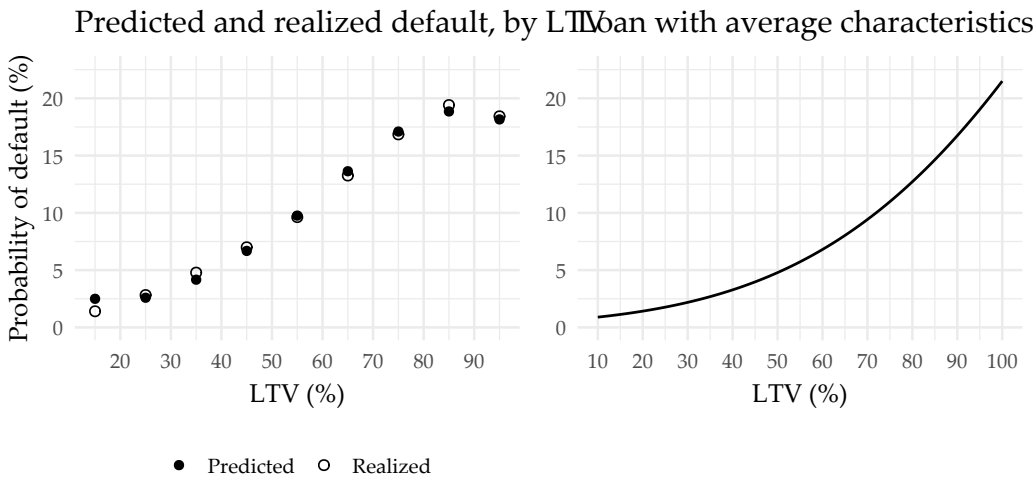
For loan i with consideration set $\mathcal{J}_i$ and winner w, let $\Omega_i \equiv \pi_i^L(L_i)R_i^*$ be the transaction rate in revenue units and $W_i \equiv \pi_i^L(L_i)\rho_i$ the (unobserved) reservation rate in the same units. Let $G_j(t) \equiv F_j^C(t-1)$, with density g_j, and F^W,f^W the distribution of W_i. By

Table D2: Share repaid among defaulted loans – baseline and with fixed effects

	(1) Baseline	(2) Fixed effects
Predicted risk	−1.3259***	−1.5188***
	(0.0433)	(0.1179)
log(Car value)	−0.1601***	−0.1875***
	(0.0139)	(0.0561)
Car spread		−0.1927***
		(0.0647)
log(Income)	−0.0107	−0.0379***
	(0.0086)	(0.0115)
Observations	41,619	36,212
Dep. var. mean	0.449	0.450
Squared correlation	0.046	0.139
RMSE	0.272	0.260
Loan term fixed effects	Yes	Yes
Origination month fixed effects	Yes	Yes
Car model fixed effects	No	Yes
Lender fixed effects	No	Yes
Seller fixed effects	No	Yes
Borrower region fixed effects	No	Yes
Borrower sex fixed effects	No	Yes
Borrower age fixed effects	No	Yes
Occupation type fixed effects	No	Yes
Risk decile fixed effects	No	Yes

Notes: Dependent variable: discounted share of the loan repaid, among loans that defaulted. Fractional logit (quasi-binomial likelihood, logit link). Column 1 is the specification in the main text. IID standard errors. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Figure D1: Probability of default by loan-to-value



Notes: Left: each loan's predicted probability of default from the probit, and realized default, averaged within LTV bins. Right: the probit at the mean of the index across loans, varying only LTV.

equation (9), $\Omega_i = \min\{1 + c_{\mathcal{J}(2)}, W_i\}$, so

$$L_i^{comp} = G_w(\Omega_i)\Big[\sum_{j\in\mathcal{J}_i\setminus w} g_j(\Omega_i) \prod_{k\in\mathcal{J}_i\setminus\{w,j\}} (1 - G_k(\Omega_i))\Big](1 - F^W(\Omega_i)), \quad |\mathcal{J}_i| > 1,$$
$$L_i^{res} = G_w(\Omega_i) \prod_{j\in\mathcal{J}_i\setminus w} (1 - G_j(\Omega_i)) f^W(\Omega_i),$$
$$L_i = \frac{L_i^{comp} + L_i^{res}}{\Pr(\text{orig}_i)}, \qquad \Pr(\text{orig}_i) = \mathbb{E}_W\Big[1 - \prod_{j\in\mathcal{J}_i} (1 - G_j(W))\Big].$$

The Jacobian from R_i^* to Ω_i, $\pi_i^L(L_i)$, does not depend on the parameters.

Costs follow Gumbel-min distributions, with a location that depends on the covariates in Table 6 and a common scale. The reservation rate in revenue units is $W_i = \exp(\theta_i) - U_0/L_i$, with $\theta_i \sim N(\mu_R, \sigma_R^2)$, a shifted log-normal that accounts for the outside option in equation (5).

Inquiries measure the consideration set with error: some lenders that inquire do not quote for the auto loan, and some lenders do not file an inquiry for loans they win. We therefore let each inquiring lender, including the winner, bid with probability $p_j = \text{logistic}(\gamma_0 + \text{logit}\,\hat{\pi}_j)$ (the coefficient on $\text{logit}\,\hat{\pi}_j$ is fixed at one), independently across lenders, which amounts to replacing G_j and g_j by $p_j G_j$ and $p_j g_j$ in the expressions above, including in the probability of origination. The first stage $\hat{\pi}_j$ is estimated from inquiry records, by lender and dealer-network status, as $\hat{\pi} = (\Pr(\text{inquiry}) - f)/(\Pr(\text{inquiry} \mid \text{win}) - f)$, where f is the probability of a recent inquiry unrelated to the auto loan, measured from each borrower's inquiries in earlier periods. Dividing by the inquiry rate among winners corrects for lenders that do not file inquiries for every loan.

D.3 Observed heterogeneity: k-means clusters

To reduce the dimension of the observed heterogeneity, we group loans into clusters using the k-means algorithm, and solve the model separately for each cluster. The clustering variables are the ones that enter borrowers' and lenders' problems: the default index at zero LTV, the lenders' cost location, car value, income, the recovery share, the size of the dealer network, and car age, all standardized separately for loans with and without a dealer network. Loans whose dealers have no network cannot receive quotes from the dealer, so we cluster them separately in proportion to their share of loans. LTV, inquiries, and default are choices or outcomes, and are therefore not used to define the clusters.

Table D3 describes the clusters used in estimation. The default rate ranges from 2.3%

Table D3: Cluster profiles ($K = 10$)

	With dealer network								No network	
	5	2	7	4	3	1	8	6	10	9
Size										
Loans	29,207	4,263	41,277	47,165	10,786	28,521	23,351	29,484	8,528	13,618
Share of loans	0.12	0.02	0.17	0.20	0.05	0.12	0.10	0.12	0.04	0.06
Clustering variables (means)										
Default index at zero LTV	−3.17	−2.79	−2.73	−2.68	−2.60	−2.51	−2.30	−1.64	−2.85	−2.24
Cost location	1.042	1.077	1.168	1.109	1.041	1.133	1.218	1.128	1.053	1.152
Car value (BRL)	37,717	55,602	27,412	41,188	90,344	33,614	22,054	35,152	56,442	32,040
Monthly income (BRL)	6,723	34,935	4,175	5,416	11,353	4,715	3,495	3,750	11,618	4,445
Recovery share	0.58	0.50	0.49	0.47	0.48	0.49	0.49	0.41	0.53	0.47
Dealer network size	5.7	5.4	6.4	6.1	5.2	2.8	4.4	5.2	0.0	0.0
Car age (years)	3.4	3.5	5.9	2.4	2.6	4.6	9.7	4.2	3.3	5.7
Outcomes (means, not used for clustering)										
Default (share)	0.023	0.089	0.098	0.093	0.082	0.123	0.179	0.342	0.071	0.189
Transacted rate (PV per BRL)	1.145	1.195	1.269	1.221	1.168	1.254	1.367	1.286	1.208	1.357
# inquiries	2.18	3.03	2.10	2.22	2.51	2.12	1.97	1.94	2.53	2.11
# inquiries, out of network	0.92	1.59	0.79	0.87	1.21	1.45	1.08	0.83	2.53	2.11
Intermediated (share)	0.63	0.61	0.62	0.65	0.64	0.64	0.63	0.65	0.01	0.02
# lenders in the market	11.8	12.0	11.8	11.9	12.1	12.1	12.0	12.0	12.2	12.3

Notes: Columns are the clusters used in estimation, ordered within stratum by the default index. Default index: probit index of default at zero LTV, including fixed effects. Cost location: fitted supply cost location of the loan. Recovery share: predicted share repaid given default. Outcomes are averaged over the cells within each cluster, weighted by loans; cells with fewer than 20 loans are not exported, which leaves 99.7% of loans. 236,200 loans.

to 34.2% across clusters, and the average car value from 22 to 90 thousand BRL. Table D4 reports the share of the variation in the clustering variables explained by the clusters. With 10 clusters, this share is 54% among loans with a dealer network and 24% among loans without one. The share increases slowly with the number of clusters, reaching 83% and 71% with 150 clusters.

D.4 GMM moments and weighting

We estimate the demand parameters by one-step GMM, with moments computed at the loan level and conditional on origination. Model moments are computed cluster by cluster, given each cluster's observed characteristics and the rate distribution implied by the supply estimates, and aggregated using the data's cluster shares. Table D5 lists the 34 moments, grouped by the feature of the data they target.

The estimator solves

$$\hat{\theta} = \arg\min_{\theta} \; (m(\theta) - \hat{m})' \, W \, (m(\theta) - \hat{m}), \qquad W_{mm} = \frac{w_{g(m)}}{|g(m)|} \frac{1}{\hat{\sigma}_m^2},$$

where $m(\theta)$ stacks the model moments, $\hat{m}$ their data counterparts, W is diagonal and fixed at the data, $g(m)$ is the group of moment m, $|g|$ the number of moments in the group, w_g

Table D4: Share of variance explained by the clusters (R^2), by number of clusters

	$K = 10$	$K = 20$	$K = 30$	$K = 50$	$K = 75$	$K = 100$	$K = 150$
With dealer network							
All variables	0.54	0.66	0.70	0.76	0.79	0.81	0.83
Default index at zero LTV	0.40	0.61	0.66	0.70	0.74	0.76	0.80
Cost location	0.65	0.73	0.76	0.82	0.84	0.86	0.88
Car value	0.58	0.67	0.72	0.77	0.81	0.82	0.84
Monthly income	0.58	0.70	0.73	0.77	0.80	0.82	0.84
Recovery share	0.55	0.65	0.70	0.76	0.79	0.81	0.83
Dealer network size	0.42	0.54	0.64	0.68	0.74	0.77	0.81
Car age	0.62	0.69	0.73	0.78	0.80	0.82	0.84
No network							
All variables	0.24	0.36	0.46	0.55	0.62	0.66	0.71
Default index at zero LTV	0.15	0.26	0.25	0.49	0.63	0.62	0.69
Cost location	0.50	0.46	0.58	0.60	0.66	0.71	0.75
Car value	0.20	0.54	0.52	0.60	0.60	0.71	0.74
Monthly income	0.19	0.34	0.38	0.63	0.63	0.63	0.73
Recovery share	0.21	0.37	0.49	0.49	0.62	0.63	0.71
Car age	0.20	0.18	0.52	0.52	0.58	0.66	0.67

Notes: Between-cluster share of the total sum of squares of the standardized clustering variables, within each stratum. The no-network stratum receives $1 + \text{round}(0.094K)$ of the K clusters, and its network size is constant. Estimation uses $K = 10$.

the group multiplier in Table D5, and $\hat{\sigma}_m^2$ the sampling variance of moment m.

The sampling variances $\hat{\sigma}_m^2$ are computed from the data. For the shares, the cross moments, the covariate interactions, and the default moments, we use the IID formula over the loans in each cell; for the default moments, the variance is computed over the loans with observed default. For the default–LTV slope and the within-cluster dispersion of LTV and inquiries, we use a parametric bootstrap, drawing each cluster's loans across LTV and inquiry cells, and default within cells, at the observed frequencies. For the intermediation moment, we use a bootstrap over clusters. For the rate moments, we use the dispersion of rates within cells. The origination target is not a sample moment, and we assign it a standard deviation of 0.02. The multipliers w_g increase the weight of origination, intermediation, and the dispersion of inquiries, and give the rate moments a small weight.

D.5 Local identification of the demand parameters

We complement the identification argument in Section 5 with two local diagnostics at the estimate. Table D6 reports the singular values of the Jacobian of the moments with respect to the parameters. All eight are positive, so the moments respond to the parameters in eight independent directions, and the parameters are locally identified. The two smallest singular values indicate directions in which the criterion is relatively flat.

Table D7 reports the sensitivity of each parameter to each group of moments, following

Table D5: Moments targeted in the demand estimation

Group	Moments	Definition	Multiplier w_g
Loan-size shares	8	$\Pr(q = q_l \mid \text{orig})$, LTV bins of 10 pp from 10–20 to 80–90 (top bin omitted)	1
Inquiry shares	5	$\Pr(N = n \mid \text{orig})$, $n = 1, \ldots, 5$ ($N \geq 6$ omitted)	1
Loan size × inquiries	1	$E[qN \mid \text{orig}]$	1
Covariate interactions	9	$z \times \{E[q], E[N], E[qN]\}$ for $z \in$ {predicted risk, log car value, log income}, standardized cluster means	1
Default	3	$E[qD]$, $E[ND]$, $E[qND]$, conditional on origination	1
Default–LTV slope	1	Within-cluster slope of $\Pr(D \mid q, k)$ on q	1
LTV dispersion	1	$\sum_k (n_k/n) \operatorname{Var}(q \mid k)$	1
Origination	1	$\Pr(\text{orig})$, external target	10^4
Intermediation	1	Loan-weighted mean absolute difference between model and data intermediation shares across clusters with a dealer network; target zero	10
Inquiry dispersion	1	Mean of $\operatorname{Var}(N \mid k, q)$ over (cluster, LTV) cells with more than 20 loans	22
Rate	1	$E[R \mid \text{orig}]$	6.8×10^{-5}
Rate × inquiries	1	$E[RN \mid \text{orig}]$	6.8×10^{-5}
Rate, single inquiry	1	$E[R \mid N = 1, \text{orig}]$	6.8×10^{-5}
Total	34		

Notes: q: LTV (bin midpoint); N: number of inquiries, top-coded at 6; D: default; R: transacted rate; k: cluster. All moments are conditional on origination, with each cluster weighted by its share of the data's loans.

Andrews, Gentzkow, and Shapiro (2017). The patterns follow the informal argument. The correlation between the taste for loan size and default, $\rho_{v,d}$, is mostly determined by the slope of default on LTV. The origination target determines the level and dispersion of the taste for loan size and the outside option. The location of the search cost depends mostly on the intermediation moment, and its dispersion on the dispersion of inquiries within cells. The dealers' arrival rate depends on the intermediation moment and on origination.

E Results

E.1 Reservation rate and participation estimates

Table E1 reports the estimates of the reservation rate distribution, estimated jointly with the costs in Table 6. In the same units as the lender costs, the median reservation rate is 20.3% a year, with a dispersion of 0.12 in logs. On average across loans, the estimated probability that the rate was set by the reservation rate, rather than by the runner-up's bid, is 0.63. The outside option is not precisely estimated, and is close to zero.

Table D6: Local identification at the demand estimate

	Value
Singular value 1	4.58e+03
Singular value 2	7.57e+02
Singular value 3	5.70e+02
Singular value 4	1.25e+02
Singular value 5	6.37e+01
Singular value 6	2.26e+01
Singular value 7	2.63e+00
Singular value 8	6.05e-01
Condition number (weighted)	7.57e+03
Condition number (unweighted)	7.73e+03

Notes: Singular values of the weighted, column-scaled Jacobian of the 34 moments with respect to the 8 estimated demand parameters. The unweighted condition number uses the column-scaled Jacobian without the weight matrix.

Table D7: Sensitivity of the demand estimates to groups of moments (%)

	Pr(LTV)	Pr(N)	E[qN]	Covariates	Default	Default slope	Var(LTV) within	Pr(orig.)	Intermed.	Var(N) within
μ_v	0.9	1.7	1.0	0.2	2.5	3.5	0.1	83.5	6.3	0.4
σ_v	2.1	2.4	2.0	0.9	1.2	7.4	4.5	67.6	11.8	0.1
$\rho_{v,d}$	8.4	3.2	0.8	2.8	11.4	51.8	7.5	11.3	2.0	0.8
U_0	11.7	2.6	3.2	1.2	1.2	3.0	4.7	65.5	6.6	0.3
Default intercept	5.9	2.5	1.0	0.9	23.4	22.7	3.5	37.7	2.3	0.2
μ_κ	7.9	9.0	5.2	2.7	3.7	5.1	11.7	18.9	29.0	6.9
σ_κ	7.2	8.7	5.4	2.1	2.7	3.7	9.1	21.7	15.3	24.1
Dealer arrivals λ	2.2	10.8	4.7	1.9	0.6	1.2	1.7	31.6	43.1	2.1

Notes: Share of each parameter's sensitivity, as in Andrews, Gentzkow, and Shapiro (2017), attributed to each group of moments: $\sum_m |\Lambda_{jm}| sd_m$ summed within group, as a percentage of the total. The three rate moments (0.0% for every parameter) are omitted. Rows sum to 100.

Table E1 also reports the participation layer. The first stage, estimated from inquiry records by lender and dealer network, varies widely across lenders: among lenders with their own estimate, its interquartile range is 0.03–0.45 outside the dealer network and 0.17–0.65 within it, and 6 of the 7 lenders with both estimates participate more often within the network. Cells in which the lender files inquiries on too few of the loans it wins use the pooled estimate (43% of lender–loan pairs), and cells with too few loans won use the lender's estimate across networks (5%). The estimated shift γ_0 lowers these probabilities, implying a mean participation probability of 0.17 across lender–loan pairs: 0.12 outside the dealer network and 0.25 within it.

Table E1: Reservation rate and participation estimates

	Estimate
Reservation rate, median (% APR)	20.27 (2.69)
Dispersion of log reservation rate, σ_R	0.118 (0.008)
Outside option (BRL), U_0	−1.47 (891.10)
Participation shift, γ_0 (logit scale)	−1.443 (0.204)
Mean participation probability, outside the dealer network	0.121
Mean participation probability, within the dealer network	0.251
Share of loans priced at the reservation rate	0.63
Loans	212,649

Notes: Maximum likelihood estimates, jointly with the cost distribution. Reservation rates scaled by expected repayment, in the same units as lender costs. The log reservation rate is normal; its median is converted to an annual rate at the average loan term and interbank rate (12.97%). Participation probability: logistic of γ_0 plus the logit of the first-stage participation, estimated from inquiry records by lender and dealer network; means over lender–loan pairs. Standard errors in parentheses, by the delta method for the median.

E.2 Rate decomposition with the reservation rate

Conditional on origination, the transacted rate is $R^* = \min\{R_{\mathcal{J}(2)}, \rho_i\}$. The decomposition in equation (14) extends to the case where the reservation rate binds:

$$
\begin{aligned}
R_i^*(L) \;=\; & \underbrace{\frac{1+c_{(1)}}{\pi_i^B}}_{\text{Mg. cost + risk adjustment}} \times \underbrace{\frac{\min\{1+c_{(2)},\, \pi_i^L \rho_i\}}{1+c_{(1)}}}_{\text{Auction market power}} \\
& \times \underbrace{\frac{\min\{1+c_{\mathcal{J}(2)},\, \pi_i^L \rho_i\}}{\min\{1+c_{(2)},\, \pi_i^L \rho_i\}}}_{\text{Search frictions}} \times \underbrace{\frac{\pi_i^B}{\pi_i^L(L)}}_{\text{Adverse selection}} ,
\end{aligned}
\tag{E1}
$$

where $c_{(k)}$ is the k-th lowest cost among all lenders in the market and $c_{\mathcal{J}(k)}$ the k-th lowest among the N_i lenders in the consideration set. When the reservation rate binds, it caps both the market-power term and the search term: search cannot lower a rate that the borrower's own reservation rate already sets.

E.3 Results by risk decile

E.4 Counterfactual surplus

Table E4 complements Table 9, reporting surplus relative to the baseline and lender profits per originated loan, both evaluated at borrowers' own default probabilities. Removing search frictions illustrates the difference between total and per-loan profits: total profits increase by 56%, as the number of loans more than doubles, while profits per loan fall by

Table E2: Baseline rate decomposition by risk decile (% APR)

						Decomposition			
Decile	Pr(orig)	Loans	Pr(D)	LTV	R^*	Cost + risk	Market power	Search	Adverse sel.
1	0.007	0.002	0.012	0.529	1.45	−3.63	4.22	0.07	0.78
2	0.023	0.005	0.018	0.540	5.84	0.67	4.22	0.19	0.75
3	0.054	0.013	0.024	0.578	9.13	3.73	4.25	0.42	0.73
4	0.084	0.021	0.031	0.633	11.88	6.09	4.30	0.77	0.72
5	0.143	0.037	0.040	0.663	14.52	8.14	4.37	1.30	0.70
6	0.224	0.060	0.054	0.675	17.33	10.22	4.45	1.99	0.68
7	0.316	0.090	0.076	0.672	20.49	12.65	4.56	2.69	0.59
8	0.578	0.180	0.115	0.671	24.62	15.77	4.73	3.88	0.24
9	0.688	0.233	0.148	0.695	28.91	18.58	4.85	5.67	−0.20
10	0.850	0.358	0.216	0.808	33.59	22.68	5.05	6.37	−0.52

Notes: Baseline equilibrium, all ten clusters. Deciles of the type-determined default probability within each cluster, 1 = safest. Pr(orig): origination probability, averaged over clusters weighted by their loans. Loans: share of originated loans. Pr(D), LTV and rates: averages over originated loans. The decomposition is the one in the counterfactual results, with the adverse-selection component $\pi_i^B/\pi_i^L(L)$ added last, each in pp of APR.

Table E3: Effects of making types public, by risk decile

	Pr(orig)		Change	LTV		R^* (APR)	
Decile	Baseline	Public	Loans (%)	Baseline	Public	Baseline	Public
1	0.007	0.009	40.2	0.529	0.553	1.45	2.47
2	0.023	0.028	37.7	0.540	0.566	5.84	7.17
3	0.054	0.065	35.3	0.578	0.607	9.13	10.72
4	0.084	0.100	32.9	0.633	0.665	11.88	13.70
5	0.143	0.169	30.3	0.663	0.716	14.52	16.53
6	0.224	0.264	27.4	0.675	0.775	17.33	19.38
7	0.316	0.361	22.0	0.672	0.841	20.49	22.21
8	0.578	0.599	6.2	0.671	0.901	24.62	25.03
9	0.688	0.710	3.6	0.695	0.935	28.91	28.78
10	0.850	0.880	12.7	0.808	0.949	33.59	33.76

Notes: Baseline equilibrium and public types (lenders observe each borrower's type), all ten clusters. Deciles of the type-determined default probability within each cluster, 1 = safest. Pr(orig): origination probability, averaged over clusters weighted by their loans. Change in loans: percentage change in the number of originated loans in the decile. LTV and rates: averages over originated loans.

Table E4: Surplus in the counterfactuals, relative to the baseline

		Surplus			Profit per loan	
	Orig.	Consumer	Lender	Total	BRL	Relative
Baseline	*1.00*	*1.00*	*1.00*	*1.00*	*2,246*	*1.00*
Adverse selection						
Public types	1.13	1.13	1.31	1.22	2,602	1.16
Search fixed	1.06	1.09	1.25	1.17	2,633	1.17
Loan size fixed	1.07	0.98	1.04	1.01	2,194	0.98
Pooling in loan size	0.98	1.21	1.14	1.17	2,598	1.16
Search frictions						
No search frictions	2.21	3.08	1.56	2.26	1,582	0.70

Notes: All columns except profit per loan in BRL are relative to the baseline. Surplus: totals over all borrowers; in the baseline, consumer surplus is 1,947 BRL, lender profit 2,246 BRL and total surplus 4,192 BRL per originated loan. Lender profit is realized profit, at borrowers' actual default probabilities. Profit per loan: realized profit per originated loan. Consumer surplus excludes the taste shocks in the choice of loan size. No adverse selection: lenders observe each borrower's type. Held at the baseline: each type keeps its baseline search effort, or its baseline choice of loan size, while the other choices adjust. Pooling in loan size: lenders hold one default belief per group of observably similar borrowers, the same at every loan size, equal to the realized default rate among the group's originated loans.

30%, as competition for each borrower increases. Removing adverse selection increases both, as borrowers take larger loans; holding loan size at the baseline, profits per loan fall slightly, by 2%.