NBER WORKING PAPER SERIES

SEARCHING FOR APPROVAL

Sumit Agarwal
John Grigsby
Ali Hortaçsu
Gregor Matvos
Amit Seru
Vincent Yao

Working Paper 27341
http://www.nber.org/papers/w27341

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2020

We thank Matteo Benetton, Peter DeMarzo, Jean-Pierre Dubé, Andreas Fuster, Jean-Francois Houde, Matt Gentzkow, Ralph Koijen, Aviv Nevo, Chris Palmer, Tarun Ramadorai, Nick Roussanov, Tobias Salz, Johannes Stroebel, Adi Sunderam, Motohiro Yogo as well as seminar participants at the 2019 American Economic Association Meetings, Banff Empirical Microeconomics Workshop, FTC Microeconomics Conference, NBER IO Meetings, University of Chicago, University of Pennsylvania, Columbia University, Northwestern University, UBC Sauder, UT Austin and Stanford GSB for useful comments. Kyeongbae Kim provided exceptional research assistance. First Version: September 2016. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Searching for Approval
Sumit Agarwal, John Grigsby, Ali Hortaçsu, Gregor Matvos, Amit Seru, and Vincent Yao
NBER Working Paper No. 27341
June 2020
JEL No. G21,G5,G51,G53,L00

ABSTRACT

We study the interaction of search and application approval in credit markets. We combine a unique dataset, which details search behavior for a large sample of mortgage borrowers, with loan application and rejection decisions. Our data reveal substantial dispersion in mortgage rates and search intensity, conditional on observables. However, in contrast to predictions of standard search models, we find a novel non-monotonic relationship between search and realized prices: borrowers, who search a lot, obtain more expensive mortgages than borrowers' with less frequent search. The evidence suggests that this occurs because lenders screen borrowers' creditworthiness, rejecting unworthy borrowers, which differentiates consumer credit markets from other search markets. Based on these insights, we build a model that combines search and screening in presence of asymmetric information. Risky borrowers internalize the probability that their application is rejected, and behave as if they had higher search costs. The model rationalizes the relationship between search, interest rates, defaults, and application rejections, and highlights the tight link between credit standards and pricing. We estimate the parameters of the model and study several counterfactuals. The model suggests that "overpayment" may be a poor proxy for consumer unsophistication since it partly represents rational search in presence of rejections. Moreover, the development of improved screening technologies from AI and big data (i.e., fintech lending) could endogenously lead to more severe adverse selection in credit markets. Finally, place based policies, such as the Community Reinvestment Act, may affect equilibrium prices through endogenous search responses rather than increased credit risk.

Sumit Agarwal
National University of Singapore
Mochtar Raidy Building
15 Kent Ridge Drive
Singapore
ushakri@yahoo.com

John Grigsby
Department of Economics
University of Chicago
1160 East 58th Street
Chicago, IL 60637
jgrigsby@uchicago.edu

Ali Hortaçsu
Kenneth C. Griffin Department of Economics
University of Chicago
1126 East 59th Street
Chicago, IL 60637
and NBER
hortacsu@uchicago.edu

Gregor Matvos
Kellogg School of Management
Northwestern University
2211 Campus Drive
Global Hub 4361
Evanston IL, 60208
and NBER
gregor.matvos@kellogg.northwestern.edu

Amit Seru
Stanford Graduate School of Business
Stanford University
655 Knight Way
and NBER
aseru@stanford.edu

Vincent Yao
Georgia State University
J. Mack Robinson College of Business
35 Broad Street NW
Atlanta, GA 30303
wyao2@gsu.edu

1 Introduction

Consumer credit markets exhibit substantial price dispersion. Borrowers with similar characteristics obtain credit with substantially different interest rates or fees in both mortgage markets (Gurun et al. 2016, Allen et al. 2014, Woodward and Hall 2012, Stroebel 2015), credit card markets (Ausubel 1991, Calem and Mester 1995, Agarwal, Chomsisengphet, Stroebel and Mahoney 2018), and auto loan markets (Argyle, Nadauld, and Palmer, 2019). A leading explanation of these facts is consumer search. Less sophisticated borrowers search less, and consequently settle for more expensive financial products. Low sophistication in search is also a frequent explanation for high cost of credit for less educated, poor, low credit score (subprime), or minority borrowers. While search is one of the primary explanations of these facts, it is rarely observed in the data. In fact, the empirical literature studying search mainly infers search behavior from the price distribution, or, in rare cases, measures search behavior from surveys, which are rarely linked to consumers' choices.

We study consumer search in the $2 trillion per year mortgage origination market using a unique and proprietary panel dataset of conforming mortgages from a large government sponsored entity (GSE) in the United States. By matching these data with consumer credit reports from a large national credit bureau, we provide a unique look at the relationship between search behavior and borrower outcomes, such as origination mortgage rates, delinquency and application acceptance decisions, conditional on a large set of observed borrower and loan characteristics.

Using this data and a quantitative model, we show that in order to understand search behavior in the mortgage market, one must acknowledge a key distinction between credit markets and markets for non-financial consumer goods: sellers in credit markets reject some borrowers because they care about borrowers' credit risk. Payoffs to sellers of most consumer products do not depend on *who* purchases their product for a given price. Credit providers' profits, on the other hand, depend directly on the probability that their customers repay their loans. As a result, creditors such as mortgage lenders, evaluate borrowers' creditworthiness, and then approve or reject customers based on this evaluation. If an application is rejected, customers must search for a mortgage with another lender. We show that incorporating this approval process differentiates search in the mortgage market from search in markets for standard goods, such as books or autos. Indeed, this approval process is not limited to the mortgage market. It is a common feature in obtaining a credit card, student and small business loans, or in auto loans, and a similar process takes place in the insurance industry, where applicants are screened for underlying risks. However, most empirical search models of do not account for this important distinction between credit and other goods.

Detailed data linking individual mortgage choices and search behavior from a large secondary market participant allows us to document the central fact in the paper. Borrowers, who search a lot, obtain higher rate mortgages than borrowers, who search little. The fact that mortgage rates, inclusive of all fees, do not decline monotonically with search is very robust. It survives across different subsamples of borrowers, after extensive controls for borrowers' characteristics using data that lenders use to set mortgage rates, and after conditioning both on location and origination date. This result cannot be generated by canonical search models such as Carlson and McAfee (1983), or those, which have been applied to the market for books (De Los Santos et al. 2012), mutual funds (Hortacsu and Syverson 2004, Roussanov et al. 2018), auto loan market (Argyle, Nadauld, and Palmer, 2019) and the mortgage market (Woodward and Hall 2012). Sorensen (2000, 2001) documents patterns in the market for prescription drugs that are consistent with common search models; the market for mortgages therefore differs in some key respects.

We document two additional facts, which show that standard search models need to be amended when applied to credit markets. First, borrowers who search are more likely to be delinquent or default on the loans ex post, even conditional on very detailed ex ante measures of their creditworthiness from the lenders' perspective, such as FICO, LTV, DTI. Second, linking approval data with search, we show a robust negative relationship between the probability of mortgage approval and the number of searches. This is the case even in a narrow time window during which prior borrower searches are not observed by the lender. Standard search models, lacking any notion of creditworthiness or application rejection, have to be altered to match the fact that borrowers who search more are more likely to fall into delinquency and have their applications rejected.

To rationalize these patterns, we develop a search model which incorporates the application screening process observed in credit markets. Borrowers search for mortgages sequentially in a market with posted prices. We depart from standard search models by letting borrowers differ in their ability to repay the loan, and assuming that this creditworthiness is private information. The correlation of creditworthiness and willingness to pay for a loan can be positive or negative, which is traditionally linked to either adverse or advantageous selection. Critically, our model captures the basic features of the institutional setting: after a mortgage application is submitted, lenders conduct an in-depth screen of the borrower to obtain an imperfect, but informative signal regarding her creditworthiness. Upon this review, the lender can either approve a mortgage, or reject the application. If the application is rejected, the borrower must search for another lender, incurring her search cost once more.

The approval process affects borrowers' search, because they account for the possibility of their application being rejected. This possibility of rejection looms larger for borrowers with low creditworthiness, because an in-depth check by the lender is likely to reveal bad information. Therefore, they are more willing to accept a high interest rate to avoid future search. In other words, because of the possibility of rejection, low creditworthiness borrowers will search as if they were financially unsophisticated, high search cost borrowers. This simple intuition has several implications.

First, one cannot infer consumers' financial sophistication, i.e. search costs, from the prices they pay. Borrowers who pay higher interest rates than other similar borrowers, such as minority borrowers, are often labeled financially unsophisticated. This intuition also arises in traditional search models: consumers with high search costs, i.e. low financial sophistication, are the ones who pay high prices. In fact, this idea is central to identification of search costs from the data (e.g. Hortacsu and Syverson (2004), Allen, Clark and Houde (2014), Roussanov et al. (2018)). This intuition has been used to shape policy geared towards low financial sophistication borrowers. Our model suggests that this inference is problematic in credit markets: consumers are willing to pay high prices as a rational response when searching in presence of rejections. Suppose minority borrowers pay higher rates than non-minority borrowers, all else equal. Our model suggests an alternative plausible explanation for such pattern: it would be a rational response of minority borrowers if they face higher rejection rates.

Second, the approval process generates endogenous adverse selection. Since low creditworthiness borrowers behave as if they have higher search costs, such borrowers are endogenously more likely to take up expensive mortgages, leading firms charging higher prices to have lower quality borrowers on average. Adverse selection arises even when creditworthy borrowers have a higher willingness-to-pay for loans, which would result in *advantageous* selection in standard frameworks. This result suggests that adverse selection could be endemic in credit and insurance markets, in which lenders screen and reject borrowers.

We estimate the model using a maximum likelihood procedure which utilizes the joint distribution of search, interest rates, default, and application approval. The estimated model successfully replicates the qualitative and quantitative patterns we observe in the data. First, frequent-searchers pay higher interest rates. These borrowers are, on average, of low unobserved creditworthiness. On the other hand, they pay high rates because of their search behavior and *not directly* due to their low creditworthiness. Their low creditworthiness implies that their mortgage applications have been rejected many times. Because the chance of future rejection is high, they are willing to accept mortgages with high interest rates. In other words, it is the high probability of rejection that induces the relationship between interest rates and search.

Second, our model can explain the relationship between search, default, and loan approvals. Because frequently rejected borrowers are likely of low creditworthiness, these frequent-searchers are more likely to default ex post. Furthermore, informative screening reveals frequent-searchers to be creditworthy less frequently than it does for infrequent-searchers. This generates the negative relationship between search and application approval that is observed in the data. Jointly, the relationship between search, interest rates, default, and application acceptance/rejection rates is consistent with the one proposed by the model.

As further validation of the mechanism proposed by the paper, we examine a population of borrowers who face almost no possibility of their mortgage application being rejected as a "placebo" test of our model. These borrowers, with approval rates of almost 98.75%, differ substantially from the overall population, whose rejection probability is approximately 18%. Our model predicts that, in the absence of any possibility of application rejection, borrowers should behave as if they were searching in any standard consumer goods market, such as the market for books. In other words, if they do not fear rejection, borrowers who search more do so to obtain cheaper mortgages, so there should be a *negative* relationship between search and realized prices. Strikingly, the data show that mortgage origination rates are monotonically decreasing in the frequency of search for the population of rarely-rejected borrowers. This stands in stark contrast to the patterns for the population at large. These results provide additional support for our model, and suggest that the non-negative relationship between search and mortgage rates for the overall population is indeed driven by the approval process rather than some other unobservable borrower characteristic.

The model estimates imply that screening is quite informative: high types are approved with a probability which is 81 percentage points higher than low types. Consistent with the existing literature on search in mortgage markets, the mean search cost is large, with each additional search being equivalent to paying an additional 29.7 basis points on a loan (bp), which equates to a cost of $30 per month for an average-sized loan in our sample, or about $1,800 if the loan is refinanced or paid off in 5 years. In addition, consumers differ in search costs, with standard deviation of 11.8 bp.

The estimated model permits several counterfactuals, which illustrate that incorporating search and rejections can be critical to analyzing policies aimed at credit markets. We first consider the impact of tightened lending standards of the sort seen during the financial crisis of 2008 (Mian and Sufi 2009, Stroebel 2015) and more recently as COVID-19 pandemic has spread.[1] Instead of modeling tightened lending standards as a contraction in supply, we model them as an increase in rejections of mortgage applications. Our model allows us to study the crisis situations mentioned above when "mortgages are cheap if you can get one." We show that lenders' reduced willingness to approve mortgages increases mortgage rates in equilibrium despite no increases in mortgage cost. Tighter lending standards also increase the impact of high search cost. In other words, keeping creditworthiness fixed, the impact is largest for the least financially sophisticated borrowers. We show that lender rejection and pricing are complements in equilibrium, resulting in substantially higher transaction and posted prices than individual responses would predict. We estimate that tighter lending standards during the crisis increased average mortgage rates by 25 bp, a substantial change of a half a standard deviation.

We next use the model to evaluate the overall equilibrium effect of lenders' ability to screen and reject credit applicants. The counterfactual illustrates that lenders' ability to screen results in a large transfer of rents from borrowers to lenders across the creditworthiness spectrum. The impact of screening is largest for low-type borrowers, who are more willing to accept expensive mortgages in face of increased rejection. This behavior by low types allows lenders to increase rates in equilibrium, spilling over to high-type borrowers who now face higher posted rates. The impact is large: removing screening from the model reduces mean realized rates by 25bp. The upper bound benefit of screening on annual bank profits is \$36 billion. This is approximately 21% of aggregate bank profits in the data.[2] Screening is therefore a key feature of the mortgage market, and significantly contributes to price variation, search costs, and bank profits. This is especially important given the growth of new screening technologies with the rise of big data and AI in financial technology, which, as Brunnermeier et al. (2020) point out, could lead banks to know borrowers' default risks better than even the borrowers themselves.

Next, we pursue two counterfactual exercises to address the question of discrimination. First, our model is suited for the analysis of a realistic redlining policy, in which a portion of lenders in the market discriminate by lowering approval rates for borrowers from the discriminated group. Such policies are increasingly a cause for concern by policymakers as they worry about AI and big data based algorithms generating such behavior by new "fintech lenders" (Fuster et al. 2020). Discrimination of this sort is more subtle, and differs from explicitly denying credit to the discriminated group, or charging different prices. We show that such policies are sustainable in a sequential search equilibrium. What's more, the redlining behavior induces borrowers from the discriminated group to pay higher interest rates on average, even if they purchase a mortgage from a lender that itself does not engage in redlining. Such discriminated groups behave as though they are financially unsophisticated, but are in fact rationally responding to the increased rejection probability into their reservation rates. Our estimates imply that if half of the lenders in a region rejected borrowers from the discriminated group at twice the rate of non-redlining lenders, average realized mortgage rates increase by 29 bp. Although this rise is concentrated amongst the discriminated group, for

[1] https://www.mpamag.com/news/fewer-people-qualify-for-mortgage-as-lenders-tighten-credit-in-march-220187.aspx [Accessed on 4/22/2020]

[2] The banking sector earned profits of \$171.3 billion in 2017 according to the FDIC. See: `https://money.cnn.com/2017/03/03/investing/bank-profits-record-high-dodd-frank/index.html`, accessed April 7, 2020

reasons already discussed, those not belonging to the discriminated group also suffer higher rates. Our findings have obvious implications for regulators as they reveal the large quantitative impact potential redlining by fintech lenders may have on the broader mortgage market.

Finally, we study the impact of "place based" policies which force or incentivize lenders to pool borrowers across creditworthiness ranges within geographic areas.[3] We consider a counterfactual exercise in which borrowers of both high and low creditworthiness in the market are accepted at the same increased rate. One might conjecture that this should increase the interest rates in the market due to increased credit risk. We find the opposite, with a mean rate decline of 27 bp. Intuitively, because of lower probability of rejection, low-type borrowers' effective search costs decrease, putting downward pressure on interest rates. The reverse is true for high-type borrowers, but the effect is smaller. The resulting equilibrium features large gains by low-types, and small losses by high-type borrowers, and lenders. Thus, place based policies like Community Revinvestment Act (CRA) move equilibrium prices significantly through endogenous search responses, which can undo the effect of increased credit risk.

Overall, our results suggest that search in credit markets differs substantially from search in other product markets. Those who search more pay higher prices for mortgages. This is due to the presence of an informative screening and rejection process. The possibility of application rejection inflates borrowers effective search costs and generates endogenous adverse selection, even if low creditworthy borrowers do not have disproportionately high willingness-to-pay for a mortgage. Accounting for this informative screening process has important consequences for the design of policies that shift credit access, such as the various place-based policies put in place over recent years.

As noted above, our paper contributes to the recent literature on price dispersion and choice frictions in the mortgage market (Gurun et al 2016, Allen et al. 2014, Woodward and Hall 2012, Alexandrov and Koulayev 2017). Allen et al. (2019) conduct a detailed study of the role played by incumbency advantage and market power on search outcomes in the Canadian mortgage market. Search frictions similarly give banks market power in our model, however our focus is on the role played by the screening/rejection mechanism on search behavior rather than the bargaining process between borrowers and lenders. Ambokar and Samaee (2019b),[4] assess the importance of search costs and market power for inaction in the mortgage refinance market. They find that search costs significantly inhibit refinancing, both directly and by indirectly giving market power to banks. Ambokar and Samaee focus on the causes and consequences of inaction in the mortgage refinance market using models in which borrowers' creditworthiness is known. In contrast, we examine the importance of informative screening and approval in the face of asymmetric information for mortgage pricing and search behavior. We illustrate that the screening and credit approval process is critical for understanding how consumers search for credit products, and more broadly, products in which the seller's payoff depends on buyer's characteristics, such as insurance.

The role played by switching costs/consumer inertia in the context of health insurance choices was studied by Handel et al. (2015). In their setting, consumers self-select into a contract from a menu of contracts, as in a number

[3]Examples are the Community Reinvestment Act (CRA), which impelled lenders in particular locations to increase their application acceptance probabilities for all borrowers, or the GSE "Declining Market" Policy of 2008.

[4]Ambokar and Samaee (2019a) incorporate mortgage search and refinancing in a New Keynesian macro model. In their model, the transmission of monetary policy is dampened relative to benchmark by reducing the benefit of refinancing for non-searchers. Much of this operates through banks' ability to statistically discriminate by charging borrowers without multiple searches a higher interest rate in the belief that they are non-searchers and thus captive shoppers.

of recent theoretical papers on the role of search frictions in environments with adverse selection (e.g. Lester et al. (2016), Guerrieri et al. (2010)). In our model, borrowers are offered only one contract, and screening is performed through a noisy technology reflecting the mortgage approval process. While the menu of contracts approach depicts many insurance markets accurately, we believe our model is a more realistic description of the mortgage market – the largest consumer credit market for households in the U.S., as well as other consumer credit markets. Finally, rational inattention has been proposed as a possible explanation for dispersion in mortgage rates, and the low take-up of beneficial refinancing opportunities (Andersen et al. 2015). Although these behavioral models provide one possible microfoundation for large search costs, they do not easily lend themselves to the direct study of search behavior, which is the focus of this paper.

More broadly, our paper links to the literature using quantitative models to study the effect of competition in financial markets (Benetton, 2018; Koijen and Yogo, 2016, 2020; Agarwal, Stroebel and Mahoney 2018; Argyle, Nadauld, and Palmer, 2019; Buchak et al. 2018, 2020, Gilbukh and Goldsmith-Pinkham, 2020, Piazzesi, Schneider, and Stroebel, 2020). Our model differs with its focus on the interaction of search and rejection. Our counterfactuals are related to a literature on how market structure alters the pass through of monetary policy to mortgage rates (Scharfstein and Sunderam 2017; Wong, 2019). Our model highlights that the search response interacts with the rejection and pricing behavior of intermediaries, shaping policy outcomes.

The remainder of the paper is organized as follows. In section 2, we describe the mortgage application process and institutional background of the mortgage market. Section 3 describes the data used in our empirical analysis. In section 4, we present the basic facts of search in mortgage markets, as well as the relationship between search and prices. We present our model of search with screening in section 5. Section 6 presents additional evidence in support of the screening mechanism central to our model, such as the relationship between search and both delinquency and approval probabilities. We describe the estimation of our model in section 7 and report its results. Finally, section 8 presents our counterfactual analyses. Section 9 concludes.

2 Credit Application Process and Inquiries

The formal process of acquiring a mortgage starts with the borrower filing an application. In the application, the borrower provides information required by the lender, such as her income, occupation, and assets. Next, the lender assesses the borrower's creditworthiness. The credit report of the borrower is "pulled" by the lender to determine the borrower's eligibility for specific loans, and the interest rate that should be charged to the borrower. This "pull" is recorded as "an inquiry" by the credit bureau. In processing the loan, the lender verifies the borrower's eligibility for loan terms. This involves verifying a borrower's income, assets and other financial information. In addition, the lender initiates an appraisal of the property, which is critical in determining the loan-to-value ratio. After gauging creditworthiness of the applicant, the lender can either approve a mortgage, or reject the application.

If the application is rejected, the borrower must search for another lender. If the application is approved, the final contract terms offered to the borrower are settled at this point. The last step involves "closing" the deal where various contractual documents are signed. The borrowers pay for the cost of obtaining their credit report, the home appraisal fee, and any loan processing costs.[5] Once the mortgage is settled, borrowers make monthly payments – either directly to the lender or to a separate loan servicer, depending on the loan.

We use the credit bureau data on "total inquiries" to capture the intensity of borrower search. We limit the search window to within 45 days of the final mortgage application , following the credit bureau definition of search. Credit bureaus entitle borrowers searching for a mortgage to a "shopping window" of 45 days during which multiple credit checks from mortgage lenders affect a borrower's credit score as if they were a single inquiry.[6] This shopping window starts with the first credit check by a mortgage lender, and applies only to credit checks from mortgage lenders and brokers; other inquiries, such as those related to credit card applications, are registered separately. As a result, borrowers are not punished for search in this 45 day window. To record searches, the credit bureau takes the last mortgage application of a borrower, and looks back 45 days prior.

Formal credit inquiries might also be triggered by lenders when consumers search for other credit products. In particular, when consumers search for credit cards or other revolving lines of credit (such as home equity line of credit or "HELOCs"), lenders also "pull" the credit score of the borrower to assess their creditworthiness. These would also be recorded as part of the "total inquiries" in the credit bureau data. We check whether non-mortgage inquiries contaminate total inquiries in two ways. One, using credit bureau data merged with approved loan information, we measure the share of mortgage-related inquiries[7] as a proportion of total inquiries for a given borrower in the one month prior to her mortgage origination. The one month window reflects the fact that data on inquiry purpose are available only from one month prior to mortgage origination. Despite the short window of one month, we find that more than 80% of total inquiries during this period are flagged as mortgage related. Given it usually takes more than one month from the original inquiry to close the mortgage, the true share is likely to be higher.

Two, we consider increases in credit limits for non-mortgage consumer credit products as possible evidence of active credit search in prior months. We focus on HELOC and credit card accounts, which also require a formal credit inquiry before approval. The instance of such credit limit changes is on average, 0% in both the month that the mortgage is originated as well as in the month preceeding origination. Notably, HELOC credit limits change by around 2% on average starting three months after mortgage origination. Similarly, credit card limits change by approximately 15% beginning two months after mortgage origination. These results provide additional evidence that consumers' search for credit cards or other unsecured credit is quite limited during the period over which we examine mortgage credit related inquiries. These observations suggest that these non-mortgage inquiries will not pollute the interpretation of total inquiries as mortgage search. This result is expected: the decision to take up a mortgage is households' largest credit decision. As a result, borrowers tend to be quite careful before applying for a mortgage. Since credit scores are adversely impacted when borrowers take up credit products, they have strong incentives not to formally search for other credit products such as credit cards before applying for a mortgage.

It is possible that borrowers search for mortgages informally without a credit pull, for example, by searching for

[5] Borrowers will usually pay between 2 and 5 percent of the purchase price in closing fees, with an average of $3,700, according to a recent Zillow survey (`https://www.zillow.com/mortgage-learning/closing-costs/`, accessed February 7, 2018).

[6] https://www.consumerfinance.gov/ask-cfpb/what-exactly-happens-when-a-mortgage-lender-checks-my-credit-en-2005/

[7] As determined by the credit bureau.

lenders and interest rates offered on the internet. However, the final terms that are offered to the borrower depend on her observable creditworthiness and value of the house. Lenders can therefore offer full contract terms only after verifying the borrower's credit score ("an inquiry") and knowing the house characteristics. Consequently, our measure appropriately captures borrower search over formal terms across lenders.

3 Data and Summary Statistics

We draw two random samples from a unique proprietary dataset obtained from a large government sponsored entity (GSE) in the United States. Our first sample contains 5.36 million applications for mortgages intended to purchase or refinance a single family property, from 2001 to 2013. The loans are originated by a variety of lenders and conform to GSE standards. We consider only loan applications with a single applicant, because they tend to have cleaner search histories at the time of application. We observe the last application by the borrower, and record all inquiries in the 45 days prior to this application. The sample contains common underwriting variables, including borrower credit score, backend debt-to-income (DTI) ratio, loan-to-value (LTV) ratio of the mortgage, mortgage contract choice, loan purpose (purchase vs refinancing), occupancy (primary residence vs investment property), application date and property location, for both approved and rejected loan applications.

Our second dataset contains approximately 1.3 million mortgages that were originated between 2001 and 2011. The shorter time period relative to application sample above reflects data sharing constraints with the GSE. At origination, we observe the borrower's credit score, the LTV ratio, the loan characteristics (origination balance and term), interest rate (inclusive of fees and points), the backend DTI ratio, whether the loan was originated through a broker, loan purpose, occupancy, and the location of the mortgaged property (zip code, MSA and state). In addition, we also have information on some of borrower's demographics, including years of school, age, gender and their monthly income at origination. The inquiries are measured 45 days prior to mortgage approval. Once the loan is originated, a servicer reports monthly performance until the end of our performance period, December 2014, or the loan terminates. A loan can terminate when the borrower chooses to prepay, or forecloses (defaults) on the property. We define default to include both foreclosures and those that have missed at least three monthly payments. The data contain mortgages originated by 175 unique lenders across the full United States.[8]

Using the social security numbers of borrowers, we merge these data with applicants' credit reports provided by a consumer credit bureau which reveal the outstanding debt balances and, crucially, the number of inquiries on the individual's file at the time of the loan application.

Table 1 reports summary statistics for our sample. Our data consist of prime borrowers. Therefore the average 725 FICO score of approved borrowers substantially exceeds that of the US population, which was 688 in April 2011,[9] The average combined loan-to-value (CLTV) ratio was 73.8% and average back-end debt-to-income ratio was 37.6. Based on observables, applicants were slightly less creditworthy, with average FICO of 707, and average CLTV of 75.3%.

[8]To limit the influence of outliers, we winsorize applications and loans lying above the 99th percentile of inquiries, interest rates, DTI, or LTV ratios.

[9]http://www.fico.com/en/blogs/risk-compliance/us-credit-quality-continues-climb-will-level/, retrieved November 11, 2016.

This difference suggests that less creditworthy borrowers face a lower probability of their mortgage applications being accepted. There is substantial heterogeneity in observed creditworthiness in our pool. The standard deviation of FICO scores is 62.5 in the loan-level dataset, and 71.6 in the application dataset. We see similarly large standard deviations in both CLTV and DTI ratios. Indeed, these loans are not without credit risk: the annualized default rate is 2.3% in our sample.

Our dataset includes loans originated through the housing boom, bust and recovery. Table 2 reports summary statistics for our two datasets across three origination periods. Almost half of our observed loan applications came before the house price peak in the fourth quarter of 2006. The other half of applications are split evenly between the crisis period (fourth quarter of 2006 through fourth quarter 2009) and the post-crisis period (2010 and later). In our loan-level sample, 43.6% were originated before the crisis, 41.7% were originated during the crisis period, and 14.7% were originated in 2010 or later. The timing difference between these two samples can be partially explained by the shorter time frame of the loan-level dataset.

4 Price Dispersion and Differences in Search

Differences in mortgage rates across borrowers have frequently been attributed to costly search.[10] However, there is little direct measurement of search behavior in this market. Here we describe the basic patterns of search in the data. We first document substantial price dispersion in the mortgage market. We then use our novel data on search to show differences in search behavior among borrowers. Last, we turn to the central fact motivating our paper: the relationship between search and mortgage rates.

4.1 Price dispersion in the mortgage market

In the mortgage market, borrowers with similar characteristics pay substantially different interest rates in the same location, and at the same point in time. This has been shown in the US subprime market (Gurun et al. 2016), as well as in Canada (Allen et al. 2014). Borrowers pay substantially different mortgage rates in our sample as well, even after adjusting these rates for points and fees. We present the full distribution of rates across three origination time periods in Figure 1A, showing substantial rate dispersion. Figure 1B presents interest rates for three different FICO based creditworthiness subsets. There is substantial mortgage rate dispersion within every subset, with interest rates differing over 3 percentage points (pp) within each group. These differences are costly. The average loan in our data is originated for $169 thousand, so each *pp* represents an additional $1,200 in interest expense every year for a 30-year fixed rate mortgage (FRM).

Differences in mortgage rates may simply reflect differences in borrowers' observables. To argue that true price dispersion exists in this market, one would ideally show that two borrowers in the same market, at the same time, with the same characteristics, paid different mortgage rates. We apply this intuition in a regression framework, and

[10] See e.g. Gurun et al 2016, Allen et al. 2014, Woodward and Hall 2012, Alexandrov and Koulayev 2017.

estimate the following specification:

$$r_{itm} = \beta X_i + \mu_t + \mu_m + \varepsilon_{itm},$$

in which r_{itm} represents the origination rate of borrower i at time t in market m. X_i are the borrower and loan characteristics, such as FICO score, LTV, DTI, income, years of education, the type of the mortgage, and whether the borrower is an investor. It is worth reiterating that we observe the actual characteristics, rather than a noisy proxy derived from borrowers' locations (e.g., years of education in a zipcode), as is used by the majority of mortgage research. In order to compare borrowers in the same market at the same point in time, we condition on state fixed effects μ_m, and on time fixed effects μ_t. Our data set was collected by the lender for the purposes of making the loan and selling it to GSEs. Thus, the controls we observe and use closely approximate the variables used to set loan rates: the R^2 from the above regression is 0.796.

The object of interest is the residual of the regression above. Mortgages with negative (positive) residuals are cheaper (more expensive) than the mean mortgage with the same characteristics. The distribution of these residuals (Figure 1C) is compressed relative to the distribution of raw origination rates, suggesting that at least some of the dispersion in rates is driven by observed borrower differences. However, a substantial amount of residual rate dispersion remains. A borrower at the 10^{th} percentile of the distribution pays an origination rate that is $0.9pp$ lower than that paid by the borrower at the 90^{th} percentile of the distribution. At the average loan amount of $169 thousand, this difference results in $1,080 larger mortgage cost per year. Our estimates of residual price dispersion of $41bp$ are similar to 50bp found in Allen et al. (2014). Meanwhile Gurun et al. (2016) find a coefficient of variation of 0.23 and 0.19 in their data on fixed- and adjustable-rate mortgages, respectively, compared with 0.15 in our data.[11] Overall, borrowers with the same characteristics, in the same market, borrowing from the same lender at the same point in time pay substantially different mortgage rates.

4.2 Borrower Search, Sophistication, and Creditworthiness

Given the large differences in mortgage rates, borrowers should have substantial incentives to search. In this section we document that different borrowers search different amounts. What's more, borrower sophistication, as proxied by their education, does not explain much variation in search. Differences in borrower creditworthiness, which do not play a role in standard search models, have substantially more success.

As we later illustrate, rejections of mortgage applications play a critical role in search. Therefore, it is important to distinguish between two groups: borrowers who apply for mortgages, and borrowers who have obtained a mortgage. The median borrower who obtains a mortgage does not search much, having only 2 inquiries on her record (Figure 2, Panel A). In fact, a borrower in the 75^{th} percentile searches 3 times. Mortgage applicants search substantially more, with a median of 9 (Panel B). This result suggests that borrowers who frequently search are less likely to be approved for a mortgage. We explore this fact more directly in Section 6.2.

[11] To test whether brand preferences or non-price aspects of a particular lender account for the observed price dispersion, we add lender×origination quarter fixed effects. Adding these increases the R^2 from 0.80 to 0.81 and reduces the residual standard deviation from $41bp$ to $40bp$.

Borrower characteristics such as education, income, age, and race have been used as proxies for consumer sophistication in the literature (Woodward and Hall 2012, Gurun et al 2016). One may argue that sophisticated consumers should have lower search costs, and therefore search more. To explore this further, consider differences in search across FICO levels in Figure 2C, and across educational attainment in Figure 2D. Consistent with intuition, the most educated borrowers search most, but the difference is slight and statistically insignificant. FICO, which measures creditworthiness, is among the strongest predictors of search: low FICO scores (below 620) search substantially more than borrowers with high FICO scores (above 720).[12] These simple facts suggest that differences in creditworthiness play an important role in understanding search in the mortgage market.

We examine whether consumer sophistication and creditworthiness proxies are correlated with search more systematically using the following regression:

$$s_{itm} = \beta X_i + \mu_{mt} + \varepsilon_{itm} \tag{1}$$

in which i indexes the mortgage applicant or borrower in market m at time t. The dependent variable s_{itm} is the number of inquiries, or an indicator that the borrower belongs to the n^{th} quartile of search, scaled by 100 for legibility. We examine the conditional correlation between search and borrower characteristics, such as their FICO score, education, income and race. To ensure that the correlation between characteristics and search is not driven by local or aggregate conditions, we include the location-time fixed effect μ_{mt}. Any differences in the regulatory environment are also absorbed by the location fixed effect. We present the results in Table 3.

Panel A reports estimates for our sample of mortgage applicants, while Panel B reports estimates for our sample of mortgage borrowers. Borrower characteristics such as education and race are correlated with the amount of search, but the simple correlations are not consistent with the intuition that sophisticated borrowers search more. More critical to our argument, more creditworthy borrowers search less, even conditional on other characteristics, suggesting an important role for creditworthiness in understanding consumer search behavior. The estimates from first column suggest that a borrower with a FICO score which is one standard deviation above the mean has 3.8 fewer inquiries on average in the application data, and 0.39 fewer inquiries in the realized loan data, conditonal on other observable characteristics. These are large magnitudes relative to the mean inquiries in both datasets (as reported in Table 1). Analysis in last four columns of both panels suggests that this pattern is driven by borrowers who search more. However, college educated borrowers, traditionally considered sophisticated, have 0.11 fewer inquiries than non-college borrowers at the time of mortgage origination.

4.3 Do Borrowers Who Search More Obtain Cheaper Mortgages?

The benchmark consumer search model suggests that search and transacted prices are negatively correlated. Though we more formally illustrate this in Section 5.5.1 the intuitive idea is as follows. Low search cost (financially savvy) consumers find searching cheap. This low search cost allows them to search more, and find cheaper products.

[12] The FICO score was designed as a measure of creditworthiness, but has also been used as a measure of consumer sophistication. If FICO proxied only for financial sophistication, one would expect the opposite: low FICO borrowers should search less, not more.

Conversely, high search cost (financially unsophisticated) consumers are willing to accept higher prices in order to avoid frequently paying their high search cost. As a result, they search less and consequently find more expensive products on average.

We first plot the average mortgage rate as a function of search for borrowers in Panel A of Figure 3. Under the benchmark model, the average price (origination rate) should monotonically decline with search. Figure 3, suggests this is not the case. As the number of searches increases from one to three, the interest rate indeed declines. However, past three inquiries, additional search is correlated with increased mortgage rates. High-inquiry borrowers, who search a lot, obtain worse (more expensive) mortgages than borrowers, in the middle of the search distribution. In the rest of this section, we present a broad array of tests to show this pattern is robust.

Figure 3 cuts the data on several other dimensions, which may drive search and mortgage pricing - FICO, race, income, and education - and plot the relationship between search and interest rates for each group. The same pattern persists for low, middle and high FICO scores, low, middle and highly educated populations, for black, white, and Hispanic borrowers, as well as for low, middle, and high income borrowers (see Appendix Figure A1). These univariate cuts of data suggest that the non-decreasing relationship between the amount of search and mortgage rates is not driven by borrower characteristics.

We next explore the relationship between mortgage rates and search in a regression framework, in which we can control for differences across markets, as well as borrower and mortgage characteristics. We estimate the following regression

$$r_{itm} = \sum_{s \geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + \gamma X_i + \varepsilon_{itm} \tag{2}$$

where i indexes the borrower who takes up a mortgage in market m at time t. The dependent variable r_{itm} is the mortgage rate. The independent variable of interest is the amount of search the borrower undertook before taking up a mortgage, s_i. The coefficients of interest β_s measure the mean change in mortgage rates for a borrower who searched s times, relative to a borrower who only searched once. To ensure that the correlation between search and mortgage rates is not driven by borrower or mortgage characteristics, we include extensive controls in X_i, such as the borrower's FICO score, loan to value ratio (LTV), investor status, product type (ARM vs FRM, purchase vs refinance), and backend DTI ratio. We also absorb any influence of local supply or demand conditions by inclduing time fixed effect μ_t and location fixed effect μ_m. These fixed effects also absorb any aggregate fluctuations, such as changes in the risk premia, or persistent differences across markets. We cluster standard errors at the state $\times$ origination quarter level. In effect, we consider two borrowers in the same location, at the same point in time, with the same observable characteristics, and compare how the interest rate charged on their mortgage differs with the amount of search.

Figure 4 plots the coefficients β_s. As the figure suggests, borrower, location, or time differences do not drive the relationship between search and interest rates. Increased search has a U-shaped, or even monotically increasing relationship with interest rates. Furthermore, these results persist if we estimate equation (2) for different borrower creditworthiness (FICO) levels, as shown in Panels B through D of Figure 4. If anything, the results are even more striking than the baseline. As in Figure 3, the low and medium FICO borrowers who search more pay the highest rates. We repeat the test in other sub-populations, which have been used to proxy for consumer sophistication or creditworthiness: race, education, and income. For brevity, we estimate a quadratic relationship between search and interest rates, rather than a fully non-parametric relationship, and present the results in Table 4. Frequent-searchers pay higher rates than borrowers who search only once, controlling for differences across borrowers, across every sub-population. This is true for low, middle and highly educated populations, for black, white, and Hispanic borrowers, as well as for low, middle, and high income borrowers.

Finally, Appendix C tests robustness of these patterns by estimating equation (2) across a number of subsamples. In particular, Appendix Figure A6 estimates equation 2 controlling for a richer set of covariates, namely the set of loan-level price adjustment (LLPA) factors used by Fannie Mae. In each case, we observe a U-shaped or positive relationship between search and interest rates in the data. Overall, the predictions from the standard search models, that more search is correlated with lower mortgage rates is rejected. We therefore develop a theory, which is able to generate these patterns.

5 Model

We now extend the standard sequential search model by adding an application approval process, which mimics the institutional features of the mortgage market described in Section 2. The model serves three primary purposes. First, with this narrow modification, the model can explain patterns of search, price, mortgage approvals, and loan performance in the data. It can explain the observed U-shaped or positive relationship between search and realized prices in the mortgage market. It also yields new testable predictions, which we test and verify in section 6. Second, it permits a deeper understanding of search in markets of asymmetric information and approvals, illustrating why adverse selection can be endemic in these markets even if preferences point to advantageous selection. Third, the model is both tractable and realistic enough to be estimated, and allows us to conduct policy-relevant counterfactual analyses in Section 8, which allow us to measure the value of screening borrowers to the intermediation sector, and evaluate the introduction of new screening technologies.

Our model is an extension of the standard sequential search model proposed by Carlson and McAfee (1983) and McCall (1970). Indeed, if all applications are accepted, the model nests this canonical model of sequential search. As in standard models, lenders post interest rates for mortgages, and borrowers search for these mortgages sequentially, incurring a constant search cost for each sampled rate. Borrowers can choose to apply for the rate draw if it is suitably attractive, or forgo the rate and continue searching by paying her search cost. However, unlike standard search models, applications are subject to approval by the lender. Upon receiving a mortgage application, lenders can perform an in-depth credit check to obtain imperfect, but informative information on the borrower's creditworthiness. The credit check is valuable, because creditworthiness is private information of the borrower, and affects the lender's profits. The lender can either approve a mortgage, or reject the application. If the application is rejected, the borrower must search for another lender.

The correlation of creditworthiness (private information) and willingness to pay for a loan can be positive or negative, which is traditionally linked to either adverse or advantageous selection. As we show, adverse selection arises in equilibrium, but it is driven by the approval process, not the standard correlation between credit quality and willingness to pay.

5.1 Setting

5.1.1 Borrowers

Consumers are indexed by iz and have two characteristics: search cost $c_i \sim G(c)$, and probability of repaying a loan in full $x_z \in \{x_h, x_l\}$, with $Pr(x_z = x_h) = \lambda$. Borrowers with high repayment ability (creditworthiness), are more likely to repay a loan than borrowers with low repayment ability: $x_h > x_l$.[13] Creditworthiness and search costs are $i.i.d$ across consumers and types.[14] A consumer iz's utility from obtaining a mortgage from lender j at rate r_j is:

$$u_{ij} = -r_j + \sigma x_z.$$

Consumers prefer loans with lower interest rates. Further, to illustrate that standard adverse/advantageous selection does not drive our results, we allow consumers with different creditworthiness to have different preferences over obtaining a mortgage. If $\sigma < 0$ then less creditworthy borrowers are more willing to take up mortgages, similar to standard adverse selection models. Conversely, if $\sigma > 0$ then more creditworthy borrowers are more willing to take up a mortgage, a feature generally attributed to advantageous selection models. As we will soon see, this parameter has no bearing on consumer search, and would only affect mortgage take-up on the extensive margin. We do not incorporate default into consumer's utility in the model: if worse consumers sort to higher interest rates, it is not because they find the option to default more valuable.[15]

5.1.2 Lenders and Mortgage Approval

Lenders post mortgage interest rates. Lenders choose from a menu of K discrete potential rates to offer, $r_k \in \{r_1, \ldots, r_K\}$.[16] Lender j's expected profit on a loan to type z at rate k is:

$$\pi_{zjk} = r_k \tilde{x}_z - m,$$

[13] We provide some empirical evidence that two types are sufficient in capturing most of the richness in the data in Appendix B

[14] The i.i.d. assumption is useful to cleanly separate the effect of search costs from creditworthiness.

[15] Indeed, there is no large difference in the relationship between search and interest rates for borrowers who default on their loans ex post compared with those who do not: re-estimating equation Equation 1 for these two populations yields similar estimates of β_s. See Appendix Figure A2.

[16] We transform the problem of choosing an offered rate may into a discrete choice problem. This assumption generates equilibrium existence in the presence of adverse selection, which can otherwise be problematic. Given that most mortgage rates (97.4% of our data) are offered in discrete 1/8pp increments this is also a reasonable approximation of the institutional environment.

in which $\tilde{x}_z$ denotes the expected repayment from a \$1 loan to a borrower with repayment ability x_z. Each lender faces a common expected cost m, which comprises the cost of capital, as well as regulatory and administrative costs.

We depart from the standard sequential search model by assuming that the potential borrower observes her creditworthiness, x_z, but the lender does not. Before obtaining a mortgage, the borrower is subject to an application approval process. The lender carries out an in-depth check of applicants' creditworthiness, which generates an informative, but imperfect, signal $s_i \in \{s_h, s_l\}$. If the borrower is of repayment ability x_z, the probability that she is revealed as a high type is $p_z = Pr\left(s_h|x_z\right)$. The in-depth review is informative $p_h \geq p_l$, so high repayment ability borrowers are more likely to be revealed as such. We assume that applications generating signal s_h (indicating the borrower is high type) are approved, while those generating s_l are rejected. We nest the benchmark model without approvals by assuming screening is uninformative, $p_h = p_l = p$.

5.2 Consumer search (Demand)

In this section we analyze how consumers search for mortgages given the distribution of rates, and the approval process used by the lenders. Let $H(\tilde{r})$ be the perceived distribution of rates offered in the market. Consumers know the distribution of offered rates $H(\tilde{r})$ in the market, but do not know which lenders offer each particular rate. As a result, consumers must search for the lowest rates in the market. Search occurs sequentially. Each period, borrower i of type z pays search cost c_i and draws a rate r from the offered rate distribution $H(\cdot)$. As is standard, draws are $i.i.d.$ with replacement. A borrower decides whether to accept the rate offer r and apply for the mortgage, or reject the offer and continue searching next period. If she applies, her application is approved with probability p_z and she drops out of the market. If, however, her application is rejected, or she chooses not to apply for the loan, she searches again.[17]

To characterize optimal search behavior, consider a consumer of type iz who was offered a mortgage with a rate r. She will keep searching as long as her cost c_i of searching is smaller than the expected gain of searching once more:

$$\begin{aligned} c_i &\leq \int_{\underline{r}}^{r} \underbrace{Pr\left(s_h|x_z\right)}_{pr.\,approval} \underbrace{\left(\left(-\tilde{r}+\sigma x_z\right)-\left(-r+\sigma x_z\right)\right)}_{better\,mortgage} dH\left(\tilde{r}\right) \\ c_i &\leq p_z \int_{\underline{r}}^{r} \left(r-\tilde{r}\right) dH\left(\tilde{r}\right) \end{aligned}$$

where $\underline{r}$ is the lowest rate offered in the market. The expected gain has two components. The first is the potential gain from finding a lower rate mortgage, $(r-\tilde{r})$. The second is the probability the borrower will be approved for the mortgage once she finds it, p_z. If borrowers are always approved $p_z = 1$, then this condition reduces to the standard search problem of Carlson and McAfee (1983). The fact that they may be rejected for a mortgage in the future increases the borrower's incentive to accept a more expensive mortgage.

Denote by r^*_{iz} the highest rate that the borrower with search cost c_i and repayment type z would accept. At this

[17] Borrowers cannot recall previously observed offered rates. Because borrowers employ a reservation price strategy, observed rates are irrelevant unless they were on rejected applications. Therefore, this assumption is equivalent to assuming that lenders will not be willing to approve a rejected borrower's future applications.

rate, the borrower is indifferent between searching further and accepting the mortgage:

$$c_i = p_z \int_{\underline{r}}^{r_{iz}^*} (r_{iz}^* - \tilde{r}) \, dH(\tilde{r}) \tag{3}$$

The borrower will optimally apply for any mortgage offered to her with interest rate less than or equal to r_{iz}^*, and will reject any mortgage offer above r_{iz}^*. Interestingly, the choice of which mortgages to accept is independent of whether there is underlying adverse or advantageous selection in the mortgage market, as σx_z drops out of the borrower's decision.[18]

As is standard in models of sequential search, reservation rates are an increasing function of search costs. From the perspective of an individual borrower, the approval process exacerbates search costs. We can see this more formally by re-writing eq. (3):

$$\frac{c_i}{p_z} = \int_{\underline{r}}^{r_{iz}^*} (r_{iz}^* - r) \, dH(r) \tag{4}$$

The search condition may therefore be rewritten into a form isomorphic to the standard search problem, in which the borrower searches with a search cost of $\frac{c_i}{p_z}$. This result also implies that without the knowledge of the approval process, one cannot infer the borrowers' search cost distribution from the price distribution alone.

5.2.1 Approval Process Induced Adverse Selection

In search markets, borrowers sort to lenders who offer different prices. The informative approval process leads to sorting on creditworthiness, resulting in adverse selection. Formally, consider two borrowers with the same search costs, but different creditworthiness. Then, from equation (3), we have:

$$p_h \int_{\underline{r}}^{r_{ih}^*} (r_{ih}^* - r) \, dH(r) = p_l \int_{\underline{r}}^{r_{il}^*} (r_{il}^* - r) \, dH(r) \, .$$

$p_h > p_l$ implies that $r_{ih}^* < r_{il}^*$. That is, less creditworthy borrowers are willing to accept higher mortgage rates than more creditworthy borrowers with the same search cost. For adverse selection to occur, the approval process must be informative. Despite underlying asymmetric information, if rejection rates are the same for both types of borrowers, $p_l = p_h$, we revert to a model with no adverse selection. Adverse selection arises even if high quality borrowers value mortgages more, i.e. if $\sigma > 0$. A positive correlation of creditworthiness and willingness to pay for a loan is traditionally linked to advantageous selection. In our model, adverse selection arises in equilibrium, but it is driven by the approval process, not the standard correlation between credit quality and willingness to pay.

To better illustrate the adverse selection problem, we present a numerical example. Figure 5A shows the distribution in reservation interest rates for high and low creditworthy types with the same normally-distributed search

[18] Note that all borrowers will continue to search until a mortgage is originated. This arises due to the implicit assumption that all borrowers find it worthwhile to take a mortgage. If borrowers instead had some outside option $\underline{u}$ to not receiving a mortgage, different values of σ may correlate with different realized shares of high and low types in the population - in essence σ may affect the equilibrium value of λ or the total market size. This paper's focus is on the search behavior of borrowers, taking as given the composition of borrowers in the market. As a result, we abstract from this consideration.

cost distribution. Creditworthy types are less willing to accept higher rates. If they find an expensive mortgage, they keep searching. Less creditworthy borrowers, on the other hand, will apply for expensive mortgages, understanding that the chances of mortgage approval are low in the future. Figure 5B shows how creditworthiness of the pool of borrowers changes as offered rates increase. Low interest rate mortgages attract borrowers of both high and low repayment ability. The market for expensive mortgages, on the other hand, is predominantly occupied by low type borrowers with high reservation rates. Differences in approval rates across types therefore lead to adverse selection.

5.3 Interest rate setting (Supply)

Lenders only accept borrowers who apply for their loan and whose credit check generates a positive signal s_h. Because borrowers sort, setting the interest rate affects both the probability of repayment on their pool of mortgages, and the expected quantity of mortgages the lender will underwrite, $S\left(\lambda q_h\left(r_j\right)+(1-\lambda) q_l(r_j)\right)$, where S is the total size of the market and $q_z(r)$ is the share of the type z market a bank charging a rate r can expect. We assume that screening is valuable, which is consistent with observing rejected applications in the mortgage market. The expected profits from charging an interest rate r are thus:[19]

$$\mathbb{E}[\Pi(r|m)] = S\left[\lambda q_h\left(r\right)\left(r\cdot\tilde{x}_h - m\right) + (1-\lambda) q_l\left(r\right)\left(r\cdot\tilde{x}_l - m\right)\right]$$

where $q_z(r)$ represents the market share of type z individuals that a bank offering rate r captures. Letting $f_z(r^*)$ be the density of reservation interest rates for borrowers of type z, Appendix D.2 shows that

$$q_z(r) = \int_r^{\infty} \frac{f_z(r^*)}{H(r^*)} dr^* \tag{5}$$

Intuitively, undirected search implies that a lender charging a rate r obtains a fraction $1/H(r^*)$ of the market for borrowers with reservation rate r^*, and can capture the mass of individuals with reservation rates above r.Note that since the function mapping offered interest rates to market shares depends on the distribution of reservation rates in the market, the distribution of search costs for participants in the market will influence banks' rate setting decision and expected profits. In essence, search costs give banks market power. This force is present in many models with search, such as the one employed by Salz (2017) to study the effects of intermediaries in the market for New York City's trade waste.

Each lender faces an additional idiosyncratic profit shock to charging specific rates $\xi_{j,k}$, which are i.i.d. and distributed according to a Type 1 Extreme Value (T1EV) distribution with scale factor σ_ξ.[20] These $\xi_{j,k}$ represent idiosyncratic lender-rate specific shocks, such as random administrative costs, the preferences of bank managers, or differences in regulatory environments. Lender k thus offers rate r_k to maximize its profits:

[19] The profit function is specified in terms of percentage points of interest. In our empirical application, we residualize observed interest rates against borrower characteristics, so that the interest rate r may take on positive or negative values. One may thus interpret Π_j as the excess return, in percentage points, that a lender may earn if it charges a rate r percentage points above the average realized rate for an observably equivalent borrower in the market.

[20] These assumptions come into play when computing counterfactuals, and do not play a role in the qualitative predictions of the model.

$$\max_{r_k \in \{r_1,\ldots,r_K\}} \mathbb{E}[\Pi(r_k|m)] + \xi_{j,k}$$

Since $\xi_{j,k}$ is i.i.d. Type 1 Extreme Value distributed with variance σ_ξ, the probability that rate r_k maximizes the lender's profit is:

$$Pr\{j \text{ choose } r_k|m, \sigma_\xi\} = \frac{\exp\left(\mathbb{E}[\Pi(r_k|m)]/\sigma_\xi\right)}{\sum_{\tilde{k}=1}^{K} \exp\left(\mathbb{E}[\Pi(r_{\tilde{k}}|m)]/\sigma_\xi\right)} \tag{6}$$

The rate setting decision outlined above will generate equilibrium price dispersion so long as σ_ξ is non-zero: the equilibrium exists despite adverse selection. Any difference in firms' cost base or regulatory environment will translate into a non-degenerate distribution of realized mortgage rates. This arises because consumer search frictions prevent the lowest-priced bank from capturing the entire market, in essence giving market power to banks.

In order to gain intuition for banks' decision, consider the impact that a unilateral small increase in the offered rate r would have on expected profits, ignoring that the rate space is in fact discrete. The derivative of the expected profit function is:

$$\frac{d\mathbb{E}[\Pi(r|m)]}{dr} = \underbrace{\underbrace{q(r)\left(\mathbb{E}\left[\tilde{x}_k|r, s_h\right]\right)}_{margin\ gain}}_{marginal\ benefit} + \underbrace{\underbrace{\frac{\partial q(r)}{\partial r}\left(r\mathbb{E}\left[\tilde{x}_k|r, s_h\right] - m\right)}_{market\ share\ loss} + \underbrace{q(r)\, r\frac{\partial \mathbb{E}\left[\tilde{x}_k|r, s_h\right]}{\partial r}}_{borrower\ pool}}_{marginal\ cost}$$

The marginal benefit of raising the mortgage rate is a higher profit on loans to existing borrowers. The marginal cost of raising prices has two components. First, the lender loses some market share $\frac{\partial q(r)}{\partial r} \leq 0$, because the marginal borrowers now choose to keep searching instead of accepting the mortgage. This downward-sloping residual demand curve highlights banks' market power in this setting (see Appendix D.2). The profits lost on each borrower are $(rE\left[\tilde{x}_k|r, s_h\right] - m) \geq 0$. The second cost of increasing mortgage rates is that a higher interest rate attracts a weakly worse pool of borrowers, $\frac{\partial E[\tilde{x}_k|r,s_h]}{\partial r} \leq 0$. The borrower pool for firms with high rates is worse because more creditworthy borrowers have lower reservation rates, and are therefore less likely to accept a mortgage when the price increases. This last component changes lenders' pricing incentives relative to a standard search model. In the benchmark model the search behavior and reservation rates are independent of borrowers' creditworthiness, which implies that $\frac{\partial E[\tilde{x}_k|r,s_h]}{\partial r} = 0$. Therefore, approvals change the lenders' pricing incentives on the margin by introducing adverse selection, which decreases incentives to raise mortgage rates on the margin.

5.4 Equilibrium

We seek pure strategy Nash equilibria. Equilibrium is defined to be an offered rate distribution $H(r)$ and a set of reservation rate strategies for high and low types $\{r_h^*(c), r_l^*(c)\}$ such that, given a set of model parameters $\{\lambda, p_h, p_l, x_h, x_l, \sigma_\xi, m\}$, and a distribution of search costs $G(c)$,

1. $H(r)$ is the distribution of optimally offered rates, chosen to maximize lender profits as in equation 6.

2. The reservation rate strategies satisfy equation 3.

3. Market shares of high and low types, $q_h(r)$ and $q_l(r)$, are calculated according to equation 5 and integrate to one; i.e.

$$\int q_z(r)dH(r) = 1 \qquad z \in l, h$$

It is important to note at this stage that the market share functions will not be degenerate. The presence of search frictions permits substantial price dispersion in equilibrium. A detailed description of our approach to computing equilibria is provided in Appendix section E.2.

5.5 Model predictions

This section presents several new predictions of our model, which differentiate it from a benchmark sequential search model in which all mortgages are approved. We test these predictions in Section 6.

5.5.1 Benchmark: All mortgages are approved

As the probability of approval for both types goes to one, the model reverts to a standard search model without the approval process.[21] We show that this benchmark model's predictions are inconsistent with the relationship between search and rates documented in Section 4.3. In this model, differences in creditworthiness are still present (i.e. $x_h \neq x_l$), and remain private information. Nevertheless, creditworthiness does not affect borrowers' search behavior: search is based solely on search costs. Substituting $p_z = 1$ into equation 3 reduces the optimal search strategy to:

$$c_i = \int_{\underline{r}}^{r^*_{iz}} (r^*_{iz} - r)\, dH(r)$$

Since high and low type individuals draw their search costs from the same distribution $G(c)$, this condition implies that both high and low type individuals have the same reservation rate distribution. As a result, there is no adverse selection despite asymmetric information – the fraction of borrowers who are high type at any particular interest rate is fixed at λ, the population share of high type borrowers.

The optimal reservation rate policy immediately makes clear that the average rate borrowers pay declines with search in equilibrium, which is the opposite of the fact we document in Section 4.3. Intuitively, the probability of an additional search is given by the probability that the borrower draws a rate higher than her reservation rate r^*_{iz}, and is thus only affected by her reservation rate: $Pr\,(Search\,again) = 1 - H\,(r^*_{iz})$. Then the probability that a borrower with a reservation rate r^* searches more than s times is:

$$Pr\,(S_{iz} > s | r^*_{iz} = r^*) = (1 - H\,(r^*))^s$$

[21] In fact, it is sufficient that $p_l = p_h = p$.

Low search cost (financially savvy) customers, have lower reservation rates, r^*, and are therefore more likely to search. Because they have lower reservation rates, their average interest rate on accepted mortgages is lower. This induces a negative relationship between search and average interest rates, as illustrated in Appendix Figure A3 for a simulated sample of borrowers. Overall, the relationship between average rates and search in the data, as shown in Section 4, rejects this prediction.

5.5.2 Informative Approval Process: Do borrowers who search more obtain cheaper mortgages?

Here we illustrate that the introduction of informative approvals can generate the non-monotonic relationship between search and transacted prices that we document in Section 4.3. The possibility of application rejection creates two reasons for a borrowers to continue to search. First, there exists the standard reason for continued search: a borrower might draw a mortgage with an interest rate above their reservation rate, $r > r^*_{iz}$, and so choose not to apply for the mortgage. Alternatively, the borrower might discover a mortgage with $r \leq r^*_{iz}$ for which they apply, only to have her application declined. The total probability that a borrower searches again is thus:

$$\begin{aligned} Pr\left(Search\, again\right) &= \underbrace{1 - Pr\left(r < r^*_{iz}\right)}_{not\, apply} + \underbrace{Pr\left(r < r^*_{iz}\right)}_{apply}\underbrace{\left(1 - p_z\right)}_{rejected} \\ &= 1 - H\left(r^*_{iz}\right) p_z. \end{aligned}$$

Therefore, the probability that a borrower with a reservation rate r^* searches more than s times is:

$$Pr\left(S_{iz} > s | r^*_{iz} = r^*\right) = \left(1 - p_z H\left(r^*\right)\right)^s$$

The two forces work in opposite directions. Less creditworthy borrowers are more willing to accept higher rates – $H(r^*_{iz})$ is higher – which pushes them to search less. However, less creditworthy borrowers are also more likely have their application rejected if they find a mortgage with a low enough rate, urging more search. If the latter force is strong enough, high type borrowers disappear from the population of searchers faster than low type borrowers. To illustrate this, we simulate a search process with highly informative screening, and plot the results in Figure 5. Panel C presents the share of high types left in the population at each level of search. With a strong screening technology, only low type individuals remain searching at the highest levels of search, while high type individuals drop out of the sample as they find acceptable mortgages.

As a result, borrowers' average reservation rate increases with the number of searches. Indeed, Figure 5D shows a positive relationship between search and realized interest rates for this simulated sample, consistent with the empirical fact documented in Section 4.3. A search model with informative applications can therefore explain the seemingly puzzling fact that borrowers, who search more, pay higher rates on average. It is worth emphasizing that rejections alone are not sufficient to explain this fact. If all borrowers are accepted with equal probability, $p_h = p_l$, the model's predictions equal that of a model without approvals, only with rescaled search costs.

5.5.3 Default and Approvals

Our model predicts a specific type of borrower sorting in equilibrium. As the number of searches increases, the creditworthiness of the borrower pool declines, as shown in Figure 5C. Defining $\tilde{\lambda}(s)$ to be the share of high type borrowers among loans realized after s inquiries, the model implies that the average default rate of borrowers with s inquiries should be $\tilde{\lambda}(s)(1-x_h)+\left(1-\tilde{\lambda}(s)\right)(1-x_l)$. Since $\tilde{\lambda}(s)$ is declining in s and $x_h > x_l$, borrowers with a large number of inquiries should be less likely to repay the lender ex post. Figure 5E illustrates the relationship between inquiries and repayment behavior for our simulated set of borrowers in our scenario with highly informative screening.

Similarly, the probability that a loan application is accepted for a borrower with s searches as $\tilde{\lambda}(s)p_h+\left(1-\tilde{\lambda}(s)\right)p_l$. Since the type of a borrower who applies for a mortgage after many searches is of lower average quality, those with high inquiry counts are more likely to be rejected upon the in-depth review. As a result, lenders are more likely to reject borrowers who search more, even if they cannot observe the number of searches. Figure 5F shows this decreasing relationship between application approval probability and inquiry counts for our simulated data. Note that in the baseline model, in which approvals are not informative, the default and approval probabilities are independent of the number of inquiries.

5.5.4 Summary

The equilibrium of our augmented search model yields the following testable predictions:

1. A non-degenerate distribution of borrower search
2. Equilibrium price dispersion in realized interest rates
3. A possibly non-monotone or non-decreasing relationship between realized interest rates and search
4. A positive relationship between search and default probability
5. A negative relationship between search and application approval probability
6. Placebo: Groups that are highly unlikely to have their application rejected (as in the benchmark model) will have a monotonically decreasing relationship between search and realized interest rates

Predictions 1 and 2 are common to search models, and are consistent with the data, as shown in Section 4. Predictions 3-5 distinguish the model with informative screening from a benchmark model without approvals. As we show in Section 4.3, the relationship between search and prices (prediction 3) is consistent with our model. We now test our model by verifying that predictions 4 through 6 are also observed in the data.

6 Additional Empirical Evidence

6.1 Loan Performance and Search

Our model predicts that less creditworthy borrowers search more in equilibrium, leading to positive relationship between search and ex-post default rates. Figure 6 plots the annualized default rate against the number of inquiries on record for all borrowers in our sample.[22] Panel A shows the annualized rate at which borrowers default, while Panel B shows the rate at which borrowers become at least 90 days delinquent on their mortgage. Both panels show that more frequent searchers are ex-post less creditworthy.

High-inquiry borrowers may simply be of lower credit quality on dimensions observable to the lender. Indeed, Figure 2C and Table 3 show that low FICO borrowers do indeed search more. In the model, borrowers who search more will be more likely to default even conditional on observables. To test this, we must take into account the fact that we only observe default behavior as of January 2015. We assume that there is a constant proportional hazard of default for all loans. Let d_{iTm} be the probability that loan i originated T years before January 2015 in market m will default in year t having survived through year $t-1$: that is, d_{itm} is the annualized hazard rate of loan i. We assume that this hazard rate has logistic form, that is:

$$d_{itm}(\theta) = \frac{\exp\left(\alpha + \sum_{s\geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_T + \mu_m + \gamma X_i + \varepsilon_{itm}\right)}{1 + \exp\left(\alpha + \sum_{s\geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_T + \mu_m + \gamma X_i + \varepsilon_{itm}\right)} \tag{7}$$

for $\theta = (\mu_m, \mu_T, \gamma, \beta_s, \alpha)$. Given this assumption, the probability that we observe that the loan has defaulted as of January 2015 is equal to one minus the probability of surviving through T years. From this we can write down the following likelihood of observing a loan's default behavior:

$$l(\text{Default}_{iTm}|\theta) = \begin{array}{cc} 1 - (1 - d_{iTm}(\theta))^T & \text{if } i \text{ defaults by Jan 2015} \\ (1 - d_{iTm}(\theta))^T & \text{if } i \text{ does not default by Jan 2015} \end{array} \tag{8}$$

We estimate the parameters θ through maximum likelihood, clustering standard errors at the origination quarter level. We define default to include both defaults and 90 day delinquency on their mortgage payments as of January 2015, scaled by 100 for legibility. The independent variable of interest is the amount of search the borrower undertook before taking up a mortgage, s_i. The coefficients of interest β_s measure the difference in log odds ratio of default for borrowers who search s times compared with those who search just once. As with our interest rate regressions, we control for observable characteristics X_i, and include a time fixed effect μ_t and location fixed effect μ_m. As before, these fixed effects absorb any aggregate fluctuations, such as changes in the regulatory environment.

We plot the coefficients of interest, β_s, in Panel C of Figure 6. Consistent with our predictions, borrowers who search more are more likely to default or become delinquent on their loans, even conditional on observable

[22] Our loan performance data is measured as of the first quarter of 2015. To generate annualized rates, we deflate the percent of mortgages which are in a state of default in January 2015 by an appropriate factor assuming a constant hazard rate and that all loans are originated at the average origination date. For instance, if $y\%$ of all loans default by January 2015 and the average loan is originated τ years before we observe loan performance, the annualized default rate $\tilde{d}$ would solve $1 - y = (1 - \tilde{d})^{\tau}$.

characteristics. The increase in log odds ratio for a borrower who searches 5 times relative to a borrower with 1 inquiry is 0.4. This translates into a borrower with 5 inquiries being approximately $e^{0.4} - 1$ =49% more likely to default on their mortgage in a given year than is a borrower with 1 inquiry, conditional on observables. This positive relationship between search and default probabilities is highly robust. We re-estimate the specification in sub-populations of low, middle and high FICO borrowers, low, middle and highly educated populations, for black, white, and Hispanic borrowers, as well as for low, middle, and high income borrowers (Panels D-F of Figure 6). Across all sub-samples, the data supports our model's prediction that more frequent searchers are on average less creditworthy than infrequent searchers, even conditional on observable characteristics.

6.2 Search and Approvals

Central to our model's predictions is the borrower approval process. The model predicts that the borrower pool of frequent searchers contains more low creditworthy types, who are more likely to be rejected following an in-depth credit check. Using our application-level dataset, we are uniquely able to test this implication of our model.[23]

Figure 7A illustrates the strong negative correlation between search and the probability of mortgage approval. This result persists in specific subsamples of our population: Figure 7A is replicated for three groups of borrower FICO scores, and across three origination time periods in Figures 7B and 7C, respectively. To illustrate that the pattern in 7 is robust, we estimate the following linear regression:

$$a_{itm} = \alpha + \sum_{s \geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + \gamma X + \varepsilon_{itm} \tag{9}$$

in which i indexes the borrower who takes up a mortgage in market m at time t. The dependent variable a_{itm} is a dummy variable taking the value 100, if the application was accepted, and 0 otherwise. Again, the coefficients of interest β_s measure the difference in acceptance probability for a borrower with s searches, compared with a borrower with just one inquiry on their credit report. As above, we include extensive controls X_i, as well as location and time fixed effects. The coefficients of interest are presented in Figure 8. Even controlling for observable loan and borrower characteristics, borrowers who search more are less likely to have their application accepted. This pattern holds across our three borrower FICO score buckets, as shown in Figure 8. The data therefore support the model's prediction that borrowers who search more more are less likely approved for mortgages, conditional on observables.

In summary, borrowers who search more are of lower average quality in two separate datasets and along two dimensions – default and application acceptance probability. The benchmark search model, in which borrowers differ only in their search cost, would predict no relationship between search and average borrower creditworthiness. This model can, therefore, neither generate the observed positive relationship between search and application rejection probability, nor the robust positive relationship between search and delinquency. What's more, the benchmark model implies that more frequent searchers pay lower interest rates on average, which is clearly rejected by the data. By contrast, our tractable model is able to generate these observed patterns in the data, both in the sample of granted

[23] Because we measure inquiries within 45 days of a mortgage application and inquiries reach a borrower's credit report with a lag, the borrower's search history is unlikely to be observed by the lender.

mortgage and among mortgage applications. We show that our model predictions hold robustly in the data, across a score of measures and subsamples.

6.3 Placebo: Borrowers who are never rejected

Our model's predictions are consistent with the the data on mortgage pricing, default, and approvals. One potential alternative explanation is that creditworthiness is observable to the lender but not the researcher, and that borrowers who search a lot are of lower creditworthiness. We think this is unlikely, since our dataset comes from lenders. Moreover, this alternative explanation does not explain why rejection rates rise with search: if creditworthiness is priced but observable, then there is no reason to reject borrowers. Nevertheless, to reject this alternative, we test another prediction from our model.

Absent the differential possibility of application rejection, our model collapses to the standard sequential search model: the borrowers who search more will, on average, borrow at lower rates. Therefore, for any subset of borrowers who do not expect to be rejected, the relationship between average rates paid and search should be negative. If, on the other hand, search is a proxy for creditworthiness observed by the lender, then we should still find a non-negative relationship, as we do for the whole sample. Intuitively, this subsample is a placebo for our proposed mechanism.

We select borrowers whose mortgage applications are rejected very rarely. We use all borrower, mortgage, location, and time characteristics to predict the probability that an application is accepted by estimating a logistic regression. Borrowers are said to be rarely-rejected if their predicted approval probability is greater than 97.5%. The average appoval rate of this sample is 98.5%. Notably, this rate is much higher than the average approval rate of 82.2% or 89.7% for high (about 720) FICO score borrowers.[24]

Panels A and B of Figure 9 show that, despite the absence of rejections, these borrowers search and face a large variation in realized mortgage rates. Indeed the search distribution for rarely-rejected borrowers is similar to that of the full population of borrowers. However the nature of this search behavior is radically different to that found in the full sample of borrowers. We plot the average mortgage origination rate of rarely rejected borrowers across searches in Figure 9C. Consistent with the model and rejecting the alternative, rarely-rejected borrowers who search more obtain mortgages with lower origination rates. This result stands in stark contrast to the positive relationship between search and mortgage rates we find for the whole population of mortgage borrowers in Figure 3. To ensure that the negative relation between search and origination rates for rarely rejected borrowers is robust, we next condition on observables by estimating regression equation 1 on this subsample. As seen in Figure 9D, rarely-rejected borrowers continue to behave as predicted by standard models of search after conditioning on observables, as our model replicates if $p_h = p_l$. These results suggest that the relationship between search, mortgage pricing, defaults, and approvals we observe is indeed driven by the informative approval process rather than some other unobservable borrower characteristic.

[24] In Appendix Figure A7 we show our results are robust to an alternative subsample of borrowers with FICO scores above 800, CLTV ratio below 60%, and a backend DTI ratio below 40%, who also attain an average approval rate of 98.5%,

7 Model Estimation and Counterfactual Analysis

Our model with search and informative approvals captures the qualitative relationship between search, mortgage rates, defaults, and approvals, which are inconsistent with standard search models. The model is rich enough to capture these patterns and is computationally tractable enough to be estimated. Estimating the model allows us to quantify the size of search costs, the underlying asymmetric information, and the value of lenders' screening technology. The estimates show the extent to which screening alters the search incentives of different types of borrowers, and the severity of the resulting adverse selection. Last, we use our estimates to study several policy relevant counterfactuals.

7.1 Estimation

The presence of multiple types presents a challenge for traditional methods of identification in search models. Because the econometrician does not directly observe market shares for high and low types separately, one is unable to recover the distribution of search costs, and approval probabilities directly from market shares and realized loan rates as in e.g. Hortaçsu and Syverson (2004). However, all parameters of the model may be identified using data on both loans and applications.

Intuitively, the difference between the distribution of search in the application and realized loan datasets identifies the application approval parameters p_z. If, for instance, all applications were approved, there would be no difference between these two search distributions. The extent to which the search distribution amongst realized loan is a parallel shift of the distribution in the application data informs the level of the approval parameters, while the differential variance and skewness of these distributions informs the gap between p_h and p_l. The relationship between default and both search and interest rates helps pin down the share of high types λ, and the default parameters x_z. Finally, the distribution of realized interest rates in the market, and the relationship between search and these interest rates informs the offered rate distribution $H(\tilde{r})$, as well as the reservation rate distributions $F_z(r^*)$, which may be inverted to recover the distribution of search costs $G(c)$. Finally, the estimated equilibrium offered rate distribution $H(\tilde{r})$ may be inverted using banks' optimal rate setting behavior to recover the banks' cost of loan origination m and the variance of the idiosyncratic profit shocks σ_ξ.

We estimate the model using maximum likelihood using our two datasets. The first dataset contains information on mortgage applications and the distribution of inquiry counts conditional on application. The second dataset is at the loan-level, and reports the origination interest rate, loan performance, and inquiry count at the time of application. That is, we observe the joint distribution of search, rates, and default, (S_i, R_i, D_i), as well as a number of observable loan and borrower characteristics. To ensure comparability of realized loans in our estimation, we residualize observed rates against observable characteristics following regression equation 2. The identification problem may be stated as follows: given the distribution of S_i conditional on application, and the joint distribution of (S_i, R_i, D_i) conditional on application approval, we must uniquely recover the set of model primitives. On the consumer side, we have to recover the search cost distribution $G(c)$, the share of creditworthy types in the population, λ, and the two repayment ability parameters, $\{x_h, x_l\}$. On the lender side, we're interested in the screening technology, $\{p_h, p_l\}$, the costs of making loans m, and the variance of the T1EV profit shocks, σ_ξ.[25]

We proceed in two steps. First, we estimate the consumer-side parameters, the screening technology parameters, and the distribution of offered rates using a maximum likelihood approach. Second, we impose that the maximum likelihood estimates of $H(\tilde{r})$ must align with the firms' choice probabilities. This suggests a robust approach to estimating the supply side parameters by minimizing the distance between our maximum likelihood estimates of $H(\tilde{r})$ and the choice probabilities as given by equation 6. We describe the construction of likelihood functions and details of our estimation in Appendix D, and computational details of the estimation in Appendix E.1. In Appendix B, we test the binary type assumption, and find little evidence for the presence of a third type.

7.2 Results

Data Fit: Despite its simplicity, the estimated model matches observed price dispersion and distribution of searches (Figure 10, Panels A and B). The model replicates an increasing relationship between interest rates and search, and search and default documented in sections 4 and 6 (Figure 10, Panels C and D).[26]

Screening Technology and Adverse Selection: Our maximum likelihood estimates are reported in Table 5. Our estimates suggest that most potential borrowers, 73%, are of low type. Assuming a constant default hazard on a 30 year mortgage, the annualized default rate of low type borrowers is $1 - 0.41^{1/30} = 2.9\%$. In expectation, low type borrowers repay 66 cents of principal on a dollar. The remaining 27% are high types, who repay almost certainly. Given that lending to a bad type is extremely costly, lenders have high incentives to screen the borrowers. Our estimates suggest that lenders make few mistakes when screening high types: p_h is close to 1, so these borrowers rarely generate a bad credit signal. That is intuitive, since a bad credit check generally requires the revelation of bad information. However, the screening process is imperfect: p_l of 19% suggests that in 19% of cases lenders' do not uncover the bad information on low types.

The difference between p_h and p_l of 0.807 suggest that the screening technology is very informative. In other words, lenders on average do a good job of verifying borrowers' income and employment, house price assessments, and other checks which are not a part of the standard measures of borrower creditworthiness such as FICO, LTV, and DTI, and reported income. A simple back of the envelope suggests that the expected loss on a bad borrower applying is lowered by approximately 81% from $34pp$ (one minus the expected repayment of a low type) to $19\% * 34pp = 6.5pp$. Therefore, given the powerful screening technology and the large benefit from successful screening, lenders find it worthwhile to screen so long as its cost is not prohibitive. As we show in Section 8.4, we estimate that the presence of this screening technology increases aggregate lender profits by \$1.2 billion per year.

The informative screening technology provides a large force towards adverse selection. Low creditworthiness

[25] Our application data reveals whether each application passed the initial approval process. This initial approval does not imply that a loan will eventually be originated, as the lender will often impose additional screening criteria after the initial approval. Thus, the approved applications in our application data do not represent the population of our loan-level data. Therefore, we do not use this application approval flag to estimate the model, and instead rely solely on the differences in the inquiry distribution in the application and loan datasets.

[26] Recall that our estimation sample consists of interest rates residualized against borrower and loan characteristics.

borrowers behave as if their search costs are $\frac{1}{19\%} = 5.3$ times higher than those of good borrowers (eq. 4), and are therefore willing to accept higher rates. To quantify the extent of adverse selection, we plot the share of borrowers at each interest rate who are expected to be high type in Figure 10E. Adverse selection is most serious for interest rates between the mean and 50bp above the mean. At the mean origination interest rate, the annualized default probability is 3.2%, and the derivative of this default rate with respect to the interest rate paid is 1.5%. Small increases in the realized interest rate lead to sizable increases in the default probability at the mean realized rate.[27] This result implies that the change in the approvals, either due to place based government policies in credit markets or technological innovation in screening can induce substantial changes in the extent of adverse selection. This, in turn, will affect the prices at which borrowers across the creditworthiness spectrum can borrow, as well as the search effort they expend in equilibrium. We examine these changes in approvals policies and technology in their full extent in Section 8.

Search Costs: The mean of the search cost distribution is estimated at 29.7 bp.[28] Our estimates of average costs are in line with 27.2bp in Allen et al. (2014), and \$29 monthly in Allen et al. (2019) for the Canadian insured mortgage market. The standard deviation of 11.8bp is smaller than 23bp in Allen et al. (2014). Furthermore, this search cost is near those estimated in the mutual fund literature, ranging from 11bp-21bp in Hortacsu and Syverson (2004) to the 39bp search cost for finding an active mutual fund in Roussanov et al. (2017).

One can translate these search costs into dollar terms using a standard mortgage calculator. Specifically, suppose a loan has origination principal Y, a term of T months, and a monthly interest rate of r (i.e. one-twelfth of the annual interest rate). The standard monthly payment for this loan is given by $y = Y\left(r(1+r)^T\right)/\left((1+r)^T - 1\right)$. This implies that the monthly payment on a 30-year fixed rate mortgage with principal of \$170,000 and interest rate of 4% per year – the mean mortgage in the data – is \$811.61. How much extra would a borrower be willing to pay in order to avoid searching one more time? If the borrower searches one additional time, she would pay the equivalent of c additional basis points of interest. Now her effective interest rate on the loan is $400 + c$ basis points. At the mean search cost of 29.7bp, this estimate would translate into a monthly payment increase of \$29.45. If a borrower moves in 5 years and prepays the mortgage, this adds up to almost \$1,800. The upper bound cost is \$10,603 over the term of the loan.[29]

Lending Cost and Margins: We estimate that the cost of making a loan, m, to be -1.59%. Because we residualize interest rates against observable characteristics before estimating the model, one should interpret m to

[27] The share of high types at each realized interest rate is analytically computed as

$$Pr\{z = h|R = r\} = \frac{Pr\{z = h \cap R = r\}}{Pr\{R = r\}} = \frac{\lambda q_h(r)}{\lambda q_h(r) + (1-\lambda)q_l(r)}$$

Likewise, the default probability of borrowers at each rate may be expressed as

$$Pr\{\text{Ever Default } |R = r\} = (1 - x_h)Pr\{z = h|R = r\} + (1 - x_l)Pr\{z = h|R = r\} = \frac{(1 - x_h)\lambda q_h(r) + (1 - x_l)(1-\lambda)q_l(r)}{\lambda q_h(r) + (1-\lambda)q_l(r)}$$

[28] As search costs are assumed to be distributed log-normally, the mean search cost is calculated as $e^{(\mu_c + \sigma_c^2/2)}$, while the standard deviation may be expressed as $\sqrt{\left(e^{\sigma_c^2} - 1\right)e^{(2\mu_c + \sigma_c^2)}}$.

[29] This estimate is an upper bound in that it assumes the mortgage is never refinanced or prepaid. In addition, we do not impose any discounting in the calculation of this upper bound cost.

be the cost of lending relative to the mean realized interest rate of a borrower with a given set of characteristics. In other words, the average markup is estimated to be 1.59%. The estimate is of the same order of magnitude as 1.09% for the insured Canadian mortgage market by Allen et al. (2014). To gauge whether these results are sensible, we can approximate the lending cost of banks as the rate on 10-year treasury bills, and compare them to the average rate on 30-year fixed rate mortgages. This average monthly spread during our sample period was 1.77%, which is very close to our estimated markup, despite the fact that we do not use any treasury rate information in our estimation.

Estimates by subsample: Finally, we estimate our model on a variety of subsamples. The results are presented in Appendix Figure A8. Estimating on observable subsamples suffers from power issues, as some subsamples do not constitute a large share of GSE loans. Nevertheless, we believe the estimates presented here to be a useful sanity check for our results. Borrower quality tends to be increasing in FICO score and decreasing in LTV ratio, whether quality is measured according to the share of high type borrowers λ (Panel A), or the repayment rates of low type borrowers x_l (Panel C). The repayment rate for high type borrowers is always 1. Search costs tend to be higher and more variables for low FICO borrowers and high LTV borrowers (Panels E and F), while screening power $p_h - p_l$ is higher for these low observable quality borrowers. Interestingly, the informativeness of the screening technology is positively correlated with the cost of misclassification $x_h - x_l$, suggesting that banks perhaps invest more in screening borrowers with low observable quality. Meanwhile, the crisis was characterized by lower repayment probabilities, lower search costs, but also less informative screening. Although these subsample estimations should be considered merely suggestive given the issues of power, we find the intuitive nature of these results to be reassuring.

8 Counterfactual Analyses

We now pursue various counterfactual exercises. We first consider the impact of tightened lending standards of the sort seen during the financial crisis of 2008 and something that is emerging again as COVID-19 pandemic has spread. Next, we estimate the impact that informative screening has on the credit market by considering a case where all mortgage applicants are accepted. We then study the practice of redlining - in which a subset of lenders selectively reject a large portion of some discriminated population. Such policies are increasingly a cause for concern by policymakers as they worry about AI and big data based algorithms generating such behavior by new "fintech lenders" (Fuster et al. 2020). We show that such practice is sustainable in a sequential search equilibrium, and induces borrowers from the discriminated group to pay higher interest rates on average. Finally, we study the impact of "place based" policies such as the Community Reinvestment Act (CRA), which direct lenders in particular locations to increase their application acceptance probabilities for all borrowers by considering a case in which screening is uninformative. The results of our counterfactual exercises are summarized in Table 6. Notably, in order to compute robust counterfactual analyses, we must recompute the distribution of equilibrium offered rates in the market. A detailed description of our approach is provided in Appendix E.2.

8.1 Tighter Lending Standards

Tightening of lending standards has been at the heart of policy debates for many years and has arisen again during the recent Covid-19 pandemic.[30] The debate has frequently centered around the trade off between providing consumers access to credit while simultaneously mitigating systematic risks in the banking sector (Dell'Ariccia et al. 2012, Mian and Sufi 2009, Bassett et al. 2014). Famously, Ben Bernanke was declined for a mortgage at the peak of the crisis during his tenure as chairman of the Federal Reserve.[31] Traditionally, a tightening in credit standards is modeled as a decline in the supply of lending because of increases in the cost of lending. Our model provides a unique opportunity to understand the implications of this tighter lending standards along new and crucial dimensions: mortgage pricing and borrowers' search response in the absence of changes in underlying costs. It allows us to speak to situations, much like in the aftermath of financial crisis of 2008 as well as the recent Covid pandemic, where "mortgages are cheap if you can get one." In other words, the costs of mortgages are low, for example, because of non-traditional monetary policy, but the probability of any individual borrower of obtaining such mortgages declines. As we show, tightening of lending standards results in higher mortgage rates even if the underlying costs of providing mortgages do not change.

In our model, a tightening of lending standards is reflected in reductions in the p_z application approval parameters. We estimate the change in approval rates during and after the crisis using a logit discrete choice model in which the dependent variable is an indicator for whether a borrower's application was approved, and controlling for observable borrower and place characteristics. We do so to adjust for changes in the pool of prospective prime borrowers (Table 2).[32] Our estimates imply a reduction of the odds ratio of application acceptance by 21.8%, suggesting that mortgage credit in fact became more difficult to attain for borrowers following the crisis. In the counterfactual, we therefore mimic the changes in mortgage approvals after the crisis by reducing the odds ratio of application approval for both high and low types by 21.8%, holding all other parameters fixed.

Even absent changes in the cost of lending or industrial structure, tightening lending standards of the magnitude seen during the crisis has quantitatively important consequences for the rates paid by borrowers. Figure 11A plots the distribution of realized rates in our tighter lending standards counterfactual (after re-computing the offered rate distribution) against the distribution implied by our baseline estimates. The mean rate paid in the market increases by 25.4bp – on the order of a discrete increment in the Fed's policy rate – and results in $301 of higher payments per year.[33] Tightening lending standards also exacerbates the distributional consequences of search costs: the standard deviation of realized interest rates increases by 9.0bp. Interestingly, tightening credit standards does not affect the extent of adverse selection in this market. Figure 11F plots the share of high types purchasing a mortgage at each rate charged in the market. The fraction of high types at each interest rate is not greatly changed, although high types become a slightly larger share of relatively high rate borrowers.

[30] https://www.mpamag.com/news/fewer-people-qualify-for-mortgage-as-lenders-tighten-credit-in-march-220187.aspx [Accessed on 4/22/2020]

[31] https://www.bloomberg.com/news/articles/2014-10-02/you-know-it-s-a-tough-market-when-ben-bernanke-can-t-refinance

[32] Specifically, we estimate a logit with state and period fixed effects for one of three periods: pre-crisis, during the crisis, and post crisis.

[33] For a 30-year fixed rate mortgage with principal of $170,000 and interest rate of 4% per year

Tightening lending standards increases the level and dispersion of equilibrium rates paid in the market through the behavior of both borrowers and lenders. The magnitudes of the results are driven by a strategic complementarity between search and pricing behavior. Broadly, the more willing borrower are to accept higher mortgage rates, the more banks can afford to charge higher prices without large reductions in market share. The intuition is the following. Borrowers will apply for a loan only if the offered rate is no higher than their reservation rate. If application rejections increase because of tightening lending standards, borrowers will be willing to accept a higher rate; i.e. their reservation rate will rise, as highlighted by equation (4). Intuitively, they fear that if they are repeatedly rejected they will have to pay multiple search costs in the future . As a result, the tightened lending standard raises borrowers' reservation rates, increasing realized rates paid by borrowers, even holding fixed the offered distribution of rates.

In response, banks react to borrowers' willingness to accept high-priced mortgages. When borrowers' reservation rates increase, banks are able to charge higher prices without sacrificing substantial market share. Intuitively, banks' residual supply curves shift out. Mathematically, as the reservation rate distribution is shifted to the right, the market share equation (5) implies that market shares will rise for every given interest rate. This incentivizes a bank to offer higher interest rates. What's more, there is a strategic complementarity – as the distribution of offered rates shifts to the right, individual banks can afford to charge higher prices without large reductions in market share. The equilibrium strategic complementarity between banks' rate setting and borrowers' reservation rates is an important driver of the magnitude of effects we find.

Overall, our model illustrates that tighter lending standards result in increased mortgage rates on the order of a discrete increment in the Fed's policy rate, as well as larger distributional consequences of search costs. A simple way to think of this is that tighter lending standards act as though credit supply is more constrained. Policies affecting credit standards must therefore take into account this direct effect on realized prices in credit markets, in addition to the standard credit access considerations.

8.2 The Value of Screening Technology

In this counterfactual, we study the overall equilibrium effect of lenders' ability to screen and reject credit applicants. This counterfactual informs us on three dimensions. First, it highlights the mechanism through which screening and rejections affect the distribution of the interest rates borrowers eventually pay, the rates posted by lenders, and costly search in equilibrium. Second, we can observe the winners and losers that emerge from lenders' improved ability to reject bad credits. This is especially important given the growth of new screening technologies with the rise of big data and AI in financial technology. Third, it allows us to compute the value lenders derive from applying screening and rejecting potential borrowers. To assess the importance of informative screening for realized rates in the market, we simulate the model under the assumption that lenders cannot screen: $p_h = p_l = 1$. This assumption also reduces our model to a benchmark model of search. We present the results in Table 6.

We find that reducing lenders' ability to screen mortgages results in a large transfer from lenders to borrowers. Without the ability to screen and reject borrowers, mortgage rates decline substantially, lowering lender profits to the benefit of borrowers. Borrowers search less, and thus realize lower cost of search. There is also a substantial redistribution between borrowers, because the distribution of offered and realized rates spreads out. Even though in equilibrium, all parties adjust simultaneously, the intuition is perhaps best understood by first focusing on low creditworthy borrowers, who are most affected by rejection. If lenders do not screen, these borrowers do not fear rejection, and therefore behave as if their effective search cost declined by a factor of 5 $\left(\frac{1}{19\%}\right)$. They are much less willing to accept a high priced mortgage. So for any distribution of offered rates, the rates that low creditworthy borrowers pay in equilibrium decline. This change in borrower behavior decreases the profitability of offering high interest rate loans. Moreover, because now high and low creditworthy borrowers are pooled, the adverse selection problem disappears. Lenders respond by lowering rates, which induces a response by high creditworthy borrowers. As we discuss above, these two forces act as strategic complements, decreasing the mean *realized* rates in the market by over 3pp. In other words, there is a large transfer from lenders to borrowers.

Given the large decrease in rates, it is not surprising that removing lenders' ability to screen leads to a substantial reduction in profits of 1.892 percentage points on every loan. This is an enormous loss: given that \$479 billion of mortgages were originated in 2017Q3,[34] the 1.892pp reduction in profits implies that the ability to rejection applications is worth approximately \$36.25 billion ($4 * 0.01892 * \$479$ billion) per year to lenders. This result implies that lenders' ability to screen is very valuable.[35] However, the result does not simply arise from screening out borrowers of low creditworthiness. As our model illustrates, with lender screening, such borrowers behave as though they have a high cost of searching, leading to a rise in lender profits.

While average rates decline, the standard deviation of realized rates in the counterfactual increases by approximately 30%. Alternatively put, screening by lenders compresses the distribution of rates in the market. Therefore, screening also results in a redistribution between borrowers. As is standard in models with adverse selection and pooling, low creditworthy types are subsidized by high creditworthy types. Moreover, because the distribution of offered rates expands without screening, the wedge between low and high search cost types increases as well. Last, removing screening decreases rejection rates, and thus reduces costly search. The average borrower searches 3.4 times under our baseline parameters, which, given a mean search cost of 29.7bp, implies a total search expense equal to approximately 101.2bp. Imposing that no borrowers are rejected reduces the mean inquiry count to 2.88, a 15% decrease in search.

This counterfactual has broad implications. If big data and AI allow lenders to screen and reject borrowers with more precision, our model implies that such innovation can lead to a large redistribution from borrowers to lenders, and between borrowers.

8.3 Discrimination and Redlining

Redlining is a practice of discrimination that denies access to products to consumers due to their socioeconomic, racial, or ethnic makeup. Such policies are increasingly a cause for concern by policymakers as they worry about

[34] https://www.newyorkfed.org/medialibrary/interactives/householdcredit/data/pdf/HHDC_2017Q3.pdf

[35] The large size of this effect, however, should be understood with the caveat that our model does not account for the entry and exit decisions of banks. Were there a fixed cost of operation, one might expect this large decline in bank profits to be met with a fewer number of banks operating, which in turn would impact both consumer credit access and the profits of operating banks.

AI and big data based algorithms generating such behavior by new "fintech lenders" (Fuster et al. 2020). There are several ways in which discrimination can occur. In Becker (1957) redlining occurs on the extensive margin: some lenders do not lend to minorities. Alternatively, discrimination can take place in prices: minority borrowers are charged higher rates. Our model is suited for the analysis of a realistic redlining policy, in which a portion of lenders in the market discriminate by lowering approval rates for borrowers from the discriminated group. Such discrimination is more subtle, and differs from explicitly denying credit to the discriminated group, or charging different prices. Incorporating realistic institutional features of the market for mortgage finance permits us to study the effect of this type of redlining on the discriminated group, as well as the consequences for the equilibrium distribution of interest rates and adverse selection in this market.[36]

We begin our redlining counterfactual by defining the discriminated group of borrowers, redlining banks, and the nature of redlining discrimination. Potential borrowers belong either to the non-discriminated group W, or the discriminated group B, the latter comprising 20% of the pool. For expositional clarity, these borrowers have identical search and creditworthiness distributions.[37] A redlining bank approves the discriminated B borrowers with 50% of the probability that the non-discriminated W borrowers of the same creditworthiness are accepted: that is, $p_z^B = 0.5p_z^W$. Half of lenders in the market redline. The non-redlining banks ignore the B, W distinction. Banks can only discriminate based on acceptance probabilities, and have to offer the same interest rates to the discriminated and non-discriminated groups. Preventing discrimination on prices focuses the mechanism on one type of redlining, and is also mostly consistent with the types of redlining and discrimination which have been alleged in this market.[38] Last, we assume that the borrowers are only aware of the proportion of banks redlining, but not *which* banks redline. This is consistent with the fact that discriminated borrowers keep applying for loans from banks, which are later alleged to have discriminated. The results are presented in Figure 12, and summarized in Table 7.

Despite the absence of discriminatory pricing, on average, discriminated borrowers, B, pay 1.6bp higher rates in equilibrium than the non-discriminated W borrowers with the same search cost and creditworthiness. The intuition is straightforward. Discriminated borrowers understand that their chances of obtaining a loan approval in the future are worse, so they are more willing to accept higher mortgage rates, *and thus* sort to banks which offer higher rates. Interestingly, the rates charged by redlining and non-redlining banks are quite similar. This is because the principal determinant of a firm's pricing decision is the distribution of reservation rates in the market; conditional on this distribution, a uniform reduction in acceptance probability does not drastically affect the firm's pricing decision on the margin.[39]

[36] Lang et al. (2005) study the effect of discrimination on markets with search by considering a model of racial bias in the labor market. In their model, black and white workers may apply to only one firm, based on a posted wage. Firms have a preference to hire white workers, despite small perceived productivity differences. As a result, black workers apply to firms where white workers are not expected to apply, realizing lower equilibrium wage rates. The intuition from their paper applies in our setting as well, however the sequential search nature of our model allows us to consider the effect of redlining on realized search costs, as well as adhere more closely to the institutional details of the mortgage market.

[37] We therefore rule out statistical discrimination, under which the discriminated characteristic would be indicative of underlying type.

[38] See, for example, Ladd (1998), and https://www.citylab.com/equity/2018/04/how-the-fair-housing-act-failed-black-homeowners/557576/, accessed Jan 17, 2019.

[39] Note that for the purposes of this counterfactual analysis, we account for the fact that redlining banks will have 50% of the market share that a non-redlining bank will have of each type of borrower when constructing the equilibrium distribution of offered rates in the market.

Becker (1957) argues that discriminating firms lose profits. In our setting, redlined lenders' profits are mostly hurt though lower volumes than the prices they charge. In order to compensate for their lower market share from disciminatory behavior, redlining banks offer slightly lower rates than do non-redlining banks on average. The mean offered rate for redlining banks is 0.291pp, as compared with 0.308pp for non-redlining banks. Similar to the prior counterfactual, because the redlined borrowers are willing to accept higher rates, banks which do not redline can charge higher rates without compromising market share. Due to the strong strategic complementarities in rate setting, redlining banks may also increase offered rates. This force leads all borrowers, not just the discriminated group, to pay higher interest rates in equilibrium, with a mean realized rate that is 28.7bp higher than in the baseline sample. However, this increase in offered rates does not offset the lost market share for the redlining banks, their profits profits decline slightly, by 2.6bp, compared with an increase in profits of 23.1bp for the non-redlining banks, relative to the baseline estimates. Put differently, redlining banks lose 25.7bp in rate of return relative to their competitors that do not engage in redlining behavior.

This counterfactual illustrates that redlining on rejections can result in the minority paying higher rates than non-minorities in equilibrium. This is despite the fact that there is no discrimination on interest rates. Discriminated minorities appear less financially sophisticated even though their underlying ability to search for mortgages equals that of the majority, i.e. they accept worse mortgages than non minorities. Our model illustrates that this is a rational response to higher expected rejection rates. Interpreting data on interest rates and rejections across potentially discriminated groups is therefore only possible in the presence of a model, which accounts for search behavior and rejections.

8.4 Place Based Policies and Pooling

Place based policies are common in mortgage and other lending markets in the U.S. These policies force or incentivize lenders to pool borrowers across creditworthiness ranges within geographic areas. In other words, lenders can in principle still reject borrowers, but cannot condition these rejections. One such policy is the Community Reinvestment Act (CRA) of 1977 requires banks to improve credit access of low socioeconomic status neighborhoods. At the same time banks are required to lend in a safe and sound manner, potentially conflicting with the first goal. An extensive literature has investigated the extent to which the CRA has increased both credit access and the riskiness of lending (Agarwal et al. 2012, Bhutta 2011, Bostic and Robinson 2003, Dahl et al. 2000). Such policies are also frequently in place in crises, such as the GSE "Declining Market" Policy of 2008. We investigate the consequences of "place based" policies such as by examining the consequences of extending credit to low creditworthiness borrowers. Our model allows us to better understand the equilibrium consequences of this policy for banks and different types of borrowers. We model an extreme version of place based polices, in which the probability of mortgage acceptance for both high and low type individuals is equalized $p_h = p_l$. This does not remove the underlying asymmetric information, just the banks' ability to act on it.[40] This stark benchmark allows us to illustrate the largest potential consequences CRA

[40] In order to maintain the same overall application acceptance probability as is observed in our data, we set $p_h = p_l = \hat{\lambda}\hat{p}_h + (1-\hat{\lambda})\hat{p}_l$, for $\hat{\lambda}, \hat{p}_h$, and $\hat{p}_l$ the estimated parameters reported in table 5.

would have in our model. The results are summarized in Table 6 and Figure 13.

One might conjecture that if banks were prevented from screening out less creditworthy borrowers, mortgage rates would increase. Instead, we find a 28.9bp decrease in average mortgage rates despite extending more credit to less creditworthy borrowers. This result broadly arises from two competing forces: because of lower probability of rejection, low-type borrowers effective search costs decrease, putting downward pressure on interest rates. High-type borrowers, on the other hand, face increased probabilities of rejection, resulting in higher effective search cost, putting upwards pressure on interest rates. The overall result is due to the large share of low-type borrowers in the market, coupled with the large increase in acceptance probabilities of low types.

Broadly, the introduction of the CRA results in a substantial redistribution between borrowers: while the average low type borrower sees reductions in realized rates of 43bp, high type individuals see rate increases of 9.1bp on average. The redistribution also occurs on non-interest rate dimensions: search, which is costly to borrowers, falls from 4.1 to 2.8 searches for low type borrowers on average, and rises from to 2.1 to 2.8 inquiries for high type borrowers. The ability to charge more to high-type borrowers blunts the cost to lenders. Lenders' profits fall by 6.3bp at the mean realized interest rate. Given a total market size of \$479billion, this 6.3bp reduction in profits implies that informative screening is worth approximately \$1.2billion ($4 * 0.00063 * \$479$ billion) per year. This loss is substantially smaller than the loss in the first counterfactual of not rejecting borrowers at all. Intuitively, when all lenders reject borrowers, they effectively impose higher search costs on the population of borrowers overall.

The intuition is that removing banks' ability to conduct in-depth credit checks also removes the resulting adverse selection, changing the pricing incentives of banks. Since low and high type individuals face the same acceptance rates, their search behavior is the same. Therefore, despite the asymmetric information problem, the adverse selection problem vanishes: at every interest rate, banks can expect a constant share of their customers to be high type.[41] The removal of the adverse selection problem effect removes an incentive for banks to shade their interest rates in order to "cream skim" high type borrowers. However, the large increase in acceptance probability for low type individuals depresses their reservation rates, putting downward pressure on the high end of the offered rate distribution. Overall, this counterfactual illustrates that accounting for the interaction between search and rejections is critical to analyzing policies which alter lenders' ability to screen and reject borrowers.

9 Conclusion

We use a novel dataset in which we observe search behavior for a large sample of mortgage borrowers. The detailed data on borrowers is matched with credit bureau data, as well as mortgage application and rejection decisions by the lenders. Consistent with search models, we find substantial dispersion in mortgage rates and search. The relationship between search and pricing that is predicted by standard search models is strongly rejected in the data: borrowers, who search a lot, obtain more expensive mortgages than borrowers, who search a moderate amount. We argue that consumer credit markets differ from other search markets because lenders use an approval process to evaluate

[41] The intuition from a standard search model dominates, so that the relationship between average prices and search becomes downward sloping, and there are flat relationships between rates, default, and search (Figure 13C).

borrowers' creditworthiness. To study how such screening influences consumer search, we develop a model of search with asymmetric information. The model predicts that search behavior is not only related to consumer sophistication, as predicted by standard search models, but also by the underlying distribution of types. The interaction between screening and search can explain why borrowers who search a lot obtain expensive mortgages, as well as account for other empirical features of the market, such as the relationship between mortgage approval and search, which standard search models cannot explain. Accounting for the credit approval process is therefore critical in understanding search behavior and equilibrium outcomes in markets for credit products, and more broadly, products in which the seller's payoff depends on buyer's characteristics, such as insurance or even labor markets.

More broadly, our paper urges that future proposals for credit market reform consider the interaction of an informative screening process with realized pricing outcomes. Such considerations present new challenges for researchers. As we show, the distribution of search costs are not identified in the presence of screening without strict data requirements. We also provide an estimation procedure for such models.

There is much scope for future research. Understanding the effect of financial education programs on mortgage market outcomes is a first order concern. Our model suggests that such programs may have little effect on equilibrium prices unless these programs also improved borrowers' creditworthiness. In addition, the fundamental economics of our model appear appropriate for a variety of settings in both consumer and producer finance, as well as in labor economics. Future research documenting whether its predictions hold in other credit markets, such as the market for credit cards, where lenders have traditionally advertised more aggressively than in mortgage markets, or the market for small business loans, where project screening may be less informative, would be very valuable.

References

[1] Agarwal, S., E. Benmelech, N. Bergman, and A. Seru (2012). Did the Community Reinvestment Act (CRA) Lead to Risky Lending? NBER Working Paper Series 18609.

[2] Agarwal, S., S. Chomsisengphet, N. Mahoney, and J. Stroebel (2018). Do Banks Pass Through Credit Expansions to Consumers Who Want To Borrow? Quarterly Journal of Economics 133(1), 129–190.

[3] Alexandrov, A. and S. Koulayev (2017). No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information. Consumer Finance Protection Bureau Office of Research Working Paper 01, 1–57.

[4] Allen, J., R. Clark, and J.-F. Houde (2014). The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry. American Economic Review 104(10), 3365–3396.

[5] Allen, J., R. Clark, and J.-F. Houde (2019). Search Frictions and Market Power in Negotiated-Price Markets. Journal of Political Economy 127(4), 1550–1598.

[6] Ambokar, S. and K. Samaee (2019a). Inaction, Search Costs, and Market Power in the US Mortgage Market. University of Pennsylvania mimeo.

[7] Ambokar, S. and K. Samaee (2019b). Mortgage Search Heterogeneity, Statistical Discrimination and Monetary Policy Transmission to Consumption. University of Pennsylvania mimeo.

[8] Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2015). Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market. NBER Working Paper Series 21386.

[9] Argyle, B., T. Nadauld, and C. J. Palmer (2017). Real Effects of Search Frictions in Consumer Credit Markets. MIT Sloan Working Paper (5242-17).

[10] Ausubel, L. M. (1991). The Failure of Competition in the Credit Card Market. American Economic Review 81(1), 50–81.

[11] Bassett, W. F., M. B. Chosak, J. C. Driscoll, and E. Zakrajsek (2014). Changes in bank lending standards and the macroeconomy. Journal of Monetary Economics 62, 23–40.

[12] Becker, G. (1957). The Economics of Discrimination (2 ed.). The University of Chicago Press.

[13] Benneton, M. (2019). Leverage Regulation and Market Structure: A Structural Model of the UK Mortgage Market. Working Paper.

[14] Bhutta, N. (2011). The Community Reinvestment Act and Mortgage Lending to Lower Income Borrowers and Neighborhoods. Journal of Law and Economics 54(4), 953–983.

[15] Bostic, R. W. and B. L. Robinson (2003). Do CRA Agreements Influence Lending Patterns? Real Estate Economics 31, 23–51.

[16] Brunnermeier, M., R. Lamba, and C. Segura-Rodriguez (2020). Inverse selection. Working Paper.

[17] Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2020). Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy. Working Paper.

[18] Buchak, G., T. Piskorski, and A. Seru (2018). Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks. Journal of Financial Economics 130(3), 453–483.

[19] Bureau, C. F. P. (2015). Consumers' mortgage shopping experience. Technical Report January.

[20] Calem, P. S. and L. J. Mester (1995). Consumer Behavior and the Stickiness of Credit-Card Interest Rates. American Economic Review 85(5), 1327–1336.

[21] Campbell, J. Y. (2016). Restoring Rational Choice: The Challenge of Consumer Financial Regulation. In Richard T Ely Lecture, Number January.

[22] Carlson, J. A. and R. P. McAfee (1983). Discrete Equilibrium Price Dispersion. Journal of Political Economy 91(3), 480–493.

[23] Dahl, D., D. D. Evanoff, and M. F. Spivey (2000). Does the Community Reinvestment Act Influence Lending? An Analysis of Changes in Bank Low-Income Mortgage Activity. Federal Reserve Bank of Chicago Working Paper.

[24] De Los Santos, B., A. Hortaçsu, and M. R. Wildenbeest (2012). Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior. American Economic Review 102(6), 2955–2980.

[25] Dell'Ariccia, G., D. Igan, and L. Laeven (2012). Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market. Journal of Money Credit and Banking 44(2), 367–384.

[26] Fuster, A., P. Goldsmith-Pinkham, T. Ramodorai, and A. Walther (2020). Predictably unequal? The effects of machine learning on credit markets. Working Paper.

[27] Gilbukh, S. and P. Goldsmith-Pinkham (2020). Heterogeneous Real Estate Agents and the Housing Cycle. Working Paper.

[28] Guerrieri, V., R. Shimer, and R. Wright (2010). Adverse Selection in Competitive Search Equilibrium. Econometrica 78(6), 1823–1862.

[29] Gurun, U. G., G. Matvos, and A. Seru (2016). Advertising Expensive Mortgages. Journal of Finance LXXI(5), 2371–2416.

[30] Handel, B., I. Hendel, and M. D. Whinston (2015). Equilibria in Health Exchanges. Econometrica 83(4), 1261–1313.

[31] Hortaçsu, A. and Syverson (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. Quarterly Journal of Economics 119(2), 403–456.

[32] Koijen, R. and M. Yogo (2016). Shadow Insurance. Econometrica 84(3), 1265–1287.

[33] Koijen, R. and M. Yogo (2020). The Fragility of Market Risk Insurance. Working Paper.

[34] Ladd, H. F. (1998). Evidence on Discrimination in Mortgage Lending. Journal of Economic Perspectives 12(2), 41–62.

[35] Lang, K., M. Manove, and W. T. Dickens (2005). Racial Discrimination in Labor Markets with Posted Wage Offers. American Economic Review 95(4), 1327–1340.

[36] Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones (2016). Screening and Adverse Selection in Frictional Markets. Federal Reserve Bank of Philadelphia Working Paper Series 16.

[37] McCall, J. J. (1970). Economics of Information and Job Search. Quarterly Journal of Economics 84(1), 113–126.

[38] Mian, A. and A. Sufi (2009). The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. Quarterly Journal of Economics 124(4), 1449–1496.

[39] Piazzesi, M., M. Schneider, and J. Stroebel (2020). Segmented Housing Search. American Economic Review 110(3).

[40] Roussanov, N., H. Ruan, and Y. Wei (2018). Marketing Mutual Funds. NBER Working Paper Series 25056.

[41] Salz, T. (2017). Intermediation and Competition in Search Markets: An Empirical Case Study. Working Paper.

[42] Scharfstein, D. and A. Sunderam (2017). Market Power in Mortgage Lending and the Transmission of Monetary Policy. Working Paper.

[43] Sorensen, A. T. (2000). Equilibrium Price Dispersion in Retail Markets for Prescription Drugs. Journal of Political Economy 108(4), 833–850.

[44] Sorensen, A. T. (2001). Price Dispersion and Heterogeneous Consumer Search for Retail Prescription Drugs. NBER Working Paper Series 8548.

[45] Stango, V. and J. Zinman (2013). Borrowing High vs Borrowing Higher: Sources and Consequences of Dispersion in Individual Borrowing Costs. NBER Working Paper Series 19069.

[46] Stroebel, J. (2015). Asymmetric Information about Collateral Values. Journal of Finance 71(3), 1071–1111.

[47] Wong, A. (2019). Refinancing and the Transmision of Monetary Policy to Consumption. Working Paper.

[48] Woodward, S. E. and R. E. Hall (2012). Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence. American Economic Review 102(7), 3249–3276.

Tables and Figures

Table 1: Summary Statistics for Mortgages and Applications

	Loan Data		Application Data	
	Mean	SD	Mean	SD
Search and Rates				
# Inquiries	2.61	2.00	12.28	11.41
Pr\{Approval\} (%)	–	–	82.19	38.26
Origination Interest Rate (%)	5.69	0.86	–	–
Creditworthiness				
FICO	725.80	62.52	707.41	71.60
CLTV	73.83	18.36	75.34	18.31
Back-end DTI ratio	37.63	12.80	37.32	12.88
Pr\{Default\} (Annualized %)	2.28	14.9	–	–
Pr\{90+ Days Delinquent\} (Annualized %)	1.40	11.7	–	–
Pr\{Prepay\} (Annualized %)	9.83	29.7	–	–
Loan Characteristics				
FRM 30-year (%)	76.60	42.33	17.86	38.30
FRM 15-year (%)	19.64	39.72	8.04	27.19
ARM (%)	3.76	19.02	74.10	43.81
Loan Origination Amount ($ 000s)	169.42	100.84	–	–
Cash-out refi (%)	31.05	46.27	0.33	5.72
Rate-term refi (%)	28.13	44.96	0.00	0.00
Borrower Characteristics				
White (%)	79.16	40.61	–	–
Black (%)	7.52	26.38	–	–
Borrower Male (%)	42.95	49.50	–	–
Borrower Age	44.44	12.52	–	–
Less than High School (%)	25.87	43.79	–	–
High School and Some College (%)	50.84	49.99	–	–
College or more (%)	18.35	38.71	–	–
Borrower Monthly Income	6104.11	6801.63	–	–
Investor (%)	8.53	27.93	8.06	27.23
Origination Date				
Pre-2006q4 (%)	43.58	49.59	48.69	49.98
2006q4–2009q4 (%)	41.68	49.30	26.59	44.18
Post-2009q4 (%)	14.74	35.45	24.72	43.14
Observations	1,316,807		5,359,060	

Notes: The first two columns report summary statistics from a sample of prime mortgages originated between January 2001 and April 2011. The latter two columns report statistics from a sample of prime mortgage applications between December 2001 and December 2013. Data provided by a large Government-Sponsored Enterprise (GSE) and merged with consumer credit reports. Payment status variables reported as of the first quarter of 2015. CLTV corresponds to combined loan-to-value ratio, while DTI stands for debt-to-income ratio.

Table 2: Average Borrower and Loan Characteristics by Time Period

Origination Date relative to 2006q4–2009q4:	Loan Data Pre	Loan Data During	Loan Data Post	Application Data Pre	Application Data During	Application Data Post
Search and Rates						
# Inquiries	1.87	3.16	3.29	14.77	11.81	7.94
Pr\{Approval\} (%)	–	–	–	80.23	82.27	85.95
Origination Interest Rate (%)	5.91	5.87	4.56	–	–	–
Creditworthiness						
FICO	713.00	726.33	762.13	689.97	701.87	747.70
CLTV	74.05	75.46	68.58	75.76	76.69	73.02
Back-end DTI ratio	36.98	39.63	33.92	37.96	39.66	33.55
Pr\{Default\} (Annualized %)	2.11	3.14	0.31	–	–	–
Pr\{90+ Days Delinquent\} (%)	1.31	1.91	0.25	–	–	–
Pr\{Prepay\} (Annualized %)	6.13	12.20	14.09	–	–	–
Loan Characteristics						
FRM 30-year (%)	71.68	85.46	66.12	18.52	9.72	25.31
FRM 15-year (%)	23.53	12.72	27.67	10.87	5.57	5.13
ARM (%)	4.78	1.83	6.20	70.61	84.71	69.56
Origination Amount ($ 000s)	138.37	187.32	210.63	–	–	–
Cash-out refi (%)	33.73	30.21	25.51	0.44	0.21	0.23
Rate-term refi (%)	26.69	25.03	41.20	0.00	0.00	0.00
Borrower Characteristics						
White (%)	80.42	77.44	80.19	–	–	–
Black (%)	8.53	8.09	2.97	–	–	–
Borrower Male (%)	44.48	41.95	40.89	–	–	–
Borrower Age	43.55	44.33	47.43	–	–	–
Less than High School (%)	26.36	27.94	18.57	–	–	–
High School and Some College (%)	46.49	53.82	55.27	–	–	–
College or more (%)	16.13	17.94	26.05	–	–	–
Borrower Monthly Income ($)	5087	6462	8095	–	–	–
Investor (%)	7.22	9.05	10.93	7.09	8.19	9.84
Observations (000s)	574	549	194	2,609	1,425	1,325

Notes: Table reports summary statistics from a sample of prime mortgages originated between January 2001 and April 2011. The first column reports statistics for loans originated before the house price peak in the fourth quarter of 2006, while column 2 reports statistics for loans originated in the crisis period between the fourth quarter of 2006 and the end of 2009. Column 3 reports statistics for loans originated in 2010 or later. Columns 4 through 6 report similar summary statistics from a sample of prime mortgage applications between December 2001 and December 2013. Data provided by a large Government-Sponsored Enterprise (GSE) and merged with consumer credit reports. Payment status variables reported as of the first quarter of 2015. CLTV corresponds to combined loan-to-value ratio, while DTI stands for debt-to-income ratio.

Table 3: Predictors of inquiry counts among mortgage applicants

	# Inquiries (1)	1^{st} Quartile (2)	2^{nd} Quartile (3)	3^{rd} Quartile (4)	4^{th} Quartile (5)
PANEL A: MORTGAGE APPLICANTS					
FICO score (std)	-3.881***	9.256***	4.831***	-1.345***	-12.743***
	(0.091)	(0.178)	(0.289)	(0.229)	(0.332)
Combined LTV (std)	0.838***	-2.853***	-0.742***	0.909***	2.686***
	(0.063)	(0.078)	(0.149)	(0.058)	(0.231)
Back-end DTI Ratio (std)	0.555***	-2.255***	-0.474***	0.914***	1.815***
	(0.018)	(0.133)	(0.086)	(0.084)	(0.072)
FRM 15-year	1.404***	-4.038***	-1.608***	0.932***	4.714***
	(0.058)	(0.194)	(0.178)	(0.153)	(0.202)
FRM 30-year	-0.495***	2.266***	0.342*	-1.008***	-1.600***
	(0.035)	(0.252)	(0.187)	(0.158)	(0.128)
Cash-out refi	-1.045***	1.099*	2.043***	0.682	-3.825***
	(0.132)	(0.593)	(0.407)	(0.444)	(0.444)
Investor	3.048***	-7.640***	-3.796***	1.516***	9.920***
	(0.142)	(0.370)	(0.364)	(0.379)	(0.458)
Observations	5202721	5202721	5202721	5202721	5202721
R^2	0.2096	0.1106	0.0190	0.0089	0.1558
PANEL B: REALIZED MORTGAGE BORROWERS					
FICO score (std)	-0.389***	6.570***	0.012***	-0.007***	-7.096***
	(0.030)	(0.271)	(0.004)	(0.002)	(0.604)
Combined LTV (std)	0.099***	-2.261***	0.000	0.004***	1.874***
	(0.007)	(0.129)	(0.001)	(0.001)	(0.160)
Back-end DTI Ratio (std)	0.120***	-2.247***	-0.003*	0.002***	2.328***
	(0.011)	(0.096)	(0.002)	(0.001)	(0.256)
FRM 15-year	-0.271***	5.266***	0.008	-0.009***	-5.156***
	(0.024)	(0.390)	(0.006)	(0.002)	(0.552)
FRM 30-year	-0.157***	3.863***	-0.002	-0.009***	-2.774***
	(0.015)	(0.576)	(0.004)	(0.002)	(0.288)
Cash-out refi	-0.141***	1.261*	0.016***	0.003*	-3.149***
	(0.040)	(0.690)	(0.003)	(0.002)	(0.842)
Black	0.270***	-4.097***	-0.007***	0.001	4.616***
	(0.019)	(0.385)	(0.002)	(0.002)	(0.330)
College	-0.109***	1.838***	0.003**	-0.002*	-1.934***
	(0.014)	(0.287)	(0.001)	(0.001)	(0.257)
Monthly Income < \$3,000	-0.173***	3.534***	0.003	-0.004***	-3.368***
	(0.017)	(0.138)	(0.003)	(0.001)	(0.395)
Investor	0.456***	-6.284***	-0.017***	-0.002	8.133***
	(0.018)	(0.575)	(0.004)	(0.003)	(0.442)
Observations	1023931	1023931	1023931	1023931	1023931
R^2	0.2378	0.2232	0.0100	0.0260	0.1731

Notes: Estimated coefficients from regression equation 1 reported. Panel A reports estimates for the sample of mortgage applications, while Panel B reports estimates for the sample of realized mortgage borrowers. Column 1 reports coefficients from a regression in which the dependent variable is the number of inquiries on an applicant's credit report. Columns 2 through 5 report coefficients from a regression in which the dependent variable is an indicator variable, scaled by 100, for whether the applicant was in the first, second, third, or fourth quartile of inquiries, respectively. Standard errors clustered at the origination quarter × state level reported in parentheses beneath coefficient. All regressions include origination quarter × state fixed effects. Coefficients marked with *, **, and *** are statistically different from 0 at the 10%, 5%, and 1% level, respectively.

Table 4: Relationship between search and origination rates within demographic groups

	Years of Education:			Monthly Income:		
	≤ 12	$13-15$	16+	$\leq$ \$3,000	\$3,001-\$7,500	> \$7,500
# Inquiries	-0.009***	-0.012***	-0.010***	-0.004	-0.008***	-0.009***
	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
# Inquiries^2	0.146***	0.181***	0.146***	0.110***	0.162***	0.161***
	(0.037)	(0.036)	(0.028)	(0.040)	(0.033)	(0.036)
Observations	327583	652322	237401	252882	748080	279436
R^2	0.7755	0.8094	0.8317	0.7478	0.7970	0.8411

	White	Black	Hispanic	Asian
# Inquiries	-0.011***	-0.002	-0.010**	-0.008**
	(0.003)	(0.004)	(0.004)	(0.003)
# Inquiries^2	0.158***	0.105**	0.166***	0.100***
	(0.031)	(0.041)	(0.042)	(0.030)
Observations	847288	77009	79678	61313
R^2	0.8078	0.7130	0.7550	0.8457

Notes: Estimated coefficients from regression equation 2 reported. Standard errors clustered at the origination quarter level reported in parentheses beneath coefficient. Dependent variable is origination interest rate plus fees and points. All regressions include lender, state, and origination quarter fixed effects, as well as controls for borrower FICO, Backend DTI ratio, CLTV, investor status, a refinance flag, and product type. Coefficients marked with *, **, and *** are statistically different from 0 at the 10%, 5%, and 1% level, respectively.

Table 5: Maximum Likelihood Estimates for our Full Sample of Loans and Applications

λ	p_h	p_l	$p_h - p_l$	x_h	x_l	$x_h - x_l$	μ_c	σ_c	μ_H	σ_H	m	σ_ξ
0.268	1.000	0.193	0.807	1.000	0.410	0.590	-1.284	0.381	0.142	0.547	-1.585	0.410
(0.002)	(0.004)	(0.000)	(0.003)	(0.003)	(0.001)	(0.004)	(0.005)	(0.004)	(0.002)	(0.001)	-	-

Notes: Table reports estimated model parameters obtained from maximum likelihood estimation described in section 7.1. Standard errors in parentheses below point estimated parameters. Parameter definitions: λ =population high type share, p_h = probability of high type application accepted, p_l = probability of low type application accepted, x_h =probability that high type repays loan in full, x_l =probability that low type repays loan in full, μ_c =mean of underlying normal distribution for log-normally distributed search costs, σ_c =standard deviation of underlying normal distribution for log-normally distributed search costs, μ_H =mean of normal distribution of equilibrium offered rates, σ_H =standard deviation of normal distribution of equilibrium offered rates, m =total bank cost of making a loan, σ_ξ =standard deviation of type-1 extreme value distributed profit shocks. The parameters governing the supply side m and σ_ξ are estimated according to the procedure outlined in Appendix Section D.3.

Table 6: Counterfactual Summary

	All Borrowers		High Type		Low Type	
	Average	S.D.	Average	S.D.	Average	S.D.
Realized Interest Rates						
Baseline MLE Estimates	0.027	0.515	-0.228	0.371	0.121	0.529
Baseline Equilibrium	-0.002	0.664	-0.384	0.459	0.140	0.673
No Rejection	-0.227	0.370	-0.228	0.371	-0.227	0.370
No Rejection (Eqm.)	-2.101	0.710	-2.103	0.709	-2.100	0.711
Tighter Lending Standards	0.035	0.523	-0.228	0.371	0.132	0.537
Tighter Lending Standards (Eqm.)	0.252	0.754	-0.217	0.497	0.427	0.759
Redlining	0.040	0.517	-0.189	0.392	0.125	0.532
Redlining (Eqm.)	0.285	0.724	0.243	0.703	0.301	0.732
Place Based Policies - CRA	0.007	0.463	0.005	0.462	0.008	0.463
Place Based Policies - CRA (Eqm.)	-0.291	0.471	-0.293	0.471	-0.290	0.471
Search Distribution						
Baseline Model Estimates	3.40	2.66	1.79	1.29	4.06	2.79
Baseline Equilibrium	3.53	2.67	2.10	1.60	4.11	2.80
No Rejection	1.79	1.29	1.79	1.29	1.79	1.29
No Rejection (w/ Supply)	2.88	2.21	2.88	2.22	2.88	2.21
Tighter Lending Standards	3.60	2.79	1.79	1.29	4.39	2.90
Tighter Lending Standards (Eqm.)	3.77	2.80	2.25	1.74	4.43	2.92
Redlining	3.45	2.70	1.77	1.26	4.14	2.82
Redlining (Eqm.)	3.24	2.74	1.12	0.46	4.12	2.81
Place Based Policies - CRA	2.76	2.08	2.76	2.08	2.76	2.07
Place Based Policies - CRA (Eqm.)	2.78	2.09	2.78	2.09	2.77	2.09

Supply Effects	Offered Rate Dist.		Bank
	Average	S.D.	Profits
Baseline MLE Estimates	0.142	0.547	1.873
Baseline Equilibrium	0.206	0.723	1.893
No Rejection (Eqm.)	-0.732	1.259	0.001
Tighter Lending Standards	0.142	0.547	1.860
Tighter Lending Standards (Eqm.)	0.483	0.805	2.130
Redlining	0.142	0.547	1.860
Redlining (Eqm.)	0.300	0.733	1.995
Place Based Policies - CRA	0.142	1.873	1.867
Place Based Policies - CRA (Eqm.)	-0.147	0.561	1.830

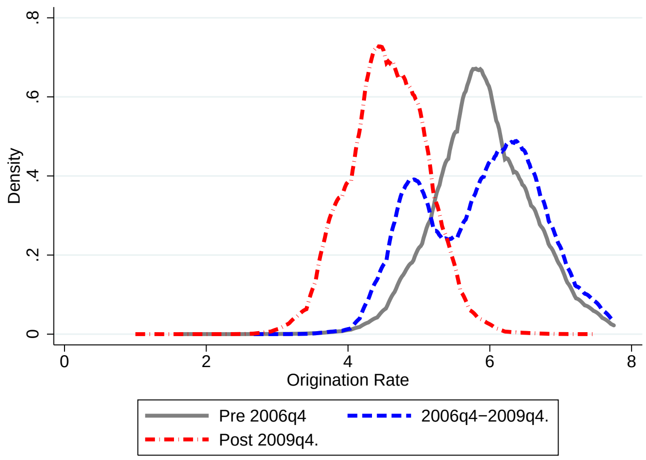
Notes: Table reports mean and standard deviation of search and realized interest rates across our counterfactual model simulations. The first two columns report mean and standard deviations for the full simulated sample of borrowers. The third and fourth columns report the mean and standard deviation for high type borrowers, while the fifth and sixth columns report the mean and standard deviation for low type borrowers. Rows with "(Eqm)" indicate counterfactual simulations in which we allow the distribution of offered rates to adjust, otherwise the offered rate distribution is fixed at those estimated in our maximum likelihood routine. Interest rates and profit margins are expressed in percentage points above the mean realized rate in the market for an observably comparable borrower and loan type. Profits reflect the profits for a bank posting the average realized rate in the market, net of any T1EV profit shocks ξ_{jk}. "No Rejection" corresponds to a model in which borrowers' applications are never rejected: $p_h = p_l = 1$. "Tighter Lending Standards" refers to a counterfactual in which the odds of application approval drop as they did following the recession, by reducing the odds of application approval by 21.8%. "Redlining" supposes that half of the lenders in the market engage in redlining behavior. For this counterfactual, borrowers belong to two groups, B or W, which are identical in all respects, except redlining banks accept applications both high and low types in the B group at half of the rate that they accept applications from the W group. We let the B group comprise 20% of the pool of borrowers. "Place Based Policies - CRA" supposes that low and high type borrowers are accepted at the same rate, maintaining the same overall acceptance probability at the estimated $\lambda p_h + (1-\lambda)p_l$.

Table 7: Redlining Counterfactual Summary

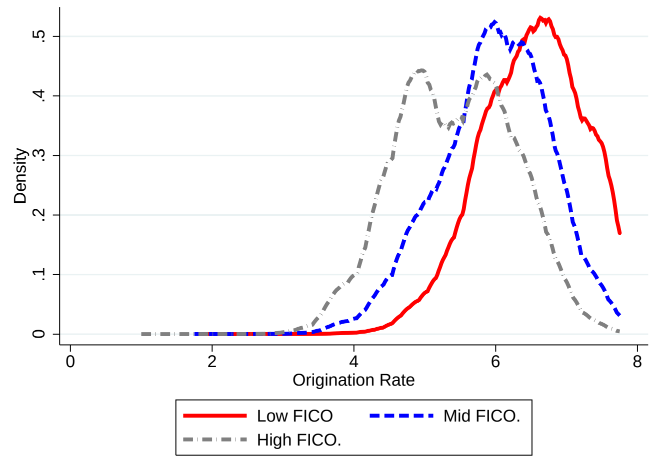
Borrower Information	Realized Rates		Search	
	Average	S.D.	Average	S.D.
Redlined Group	0.298	0.723	3.54	2.89
Non-Redlined Group	0.282	0.725	3.17	2.70
Bank Information	Offered Rates		Expected Profits	
	Average	S.D.		
Redlining Banks	0.291	0.753	1.867	
Non-Redlining Banks	0.308	0.713	2.124	

Notes: Table reports mean and standard deviation of search and realized interest rates for the group of redlined borrowers B and borrowers not subject to redlining, W. To construct this counterfactual, we suppose that redlined borrowers B comprise 20% of the borrower pool and are accepted at half the rate of W borrowers if they meet a redlining bank. Half of all banks redline. The first panel reports the effect of redlining on borrowers, while the second panel reports the effect of redlining on banks. The first two columns of the first panel report the mean and standard deviation of realized mortgage interest rates for a simulated set of borrowers, while the third and fourth columns report the mean and standard deviation of total search for these borrowers. Meanwhile, the first two columns of the second panel reports the mean and standard deviation of the offered rate distribution for redlining and non-redlining banks, estimated according to the routine outlined in Appendix E.2. The third column represents the profits of a bank charging the mean realized rate in the economy. Interest rates and profit margins are expressed in percentage points above the mean realized rate in the market for an observably comparable borrower and loan type. Profits reflect the profits for a bank posting the average realized rate in the market, net of any T1EV profit shocks ξ_{jk}.

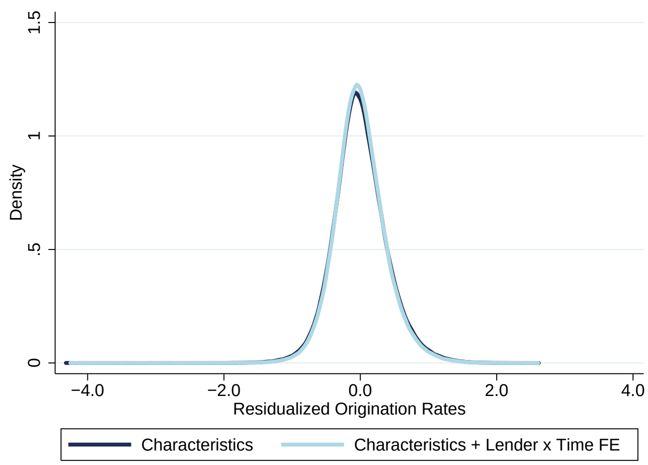
Figure 1: Distribution of Mortgage Rates in the U.S.



Panel A: Raw Rates by Origination Date



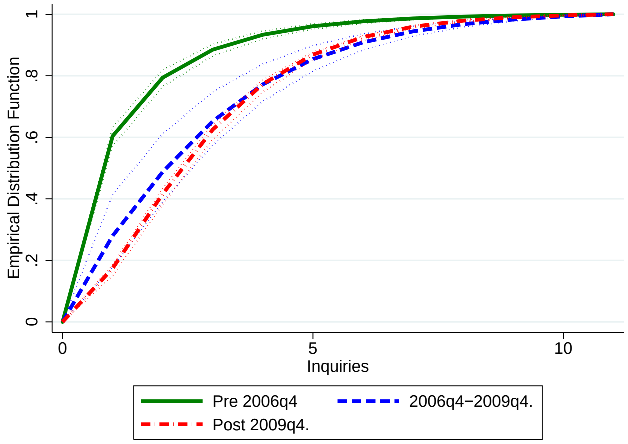
Panel B: Raw Rates by Borrower FICO Score



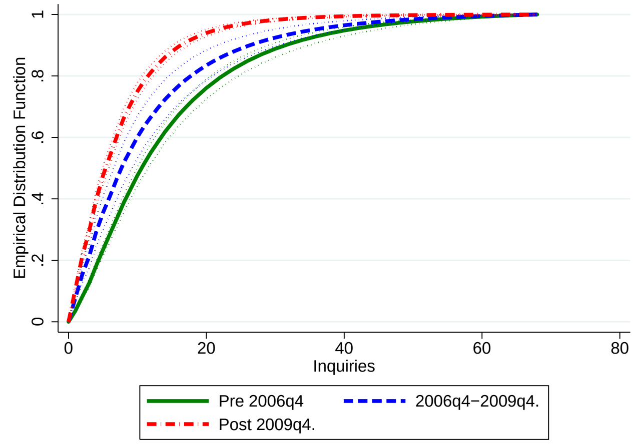
Panel C: Rates Residualized Against Observables

Notes: Figure plots the kernel-density estimated distribution of mortgage rates in the U.S. Panel A plots the raw observed rates across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post crisis period from the first quarter of 2010 on. Panel B plots the distribution of observed mortgage rates for three borrower FICO buckets: low FICO ($\leq$ 620), middle FICO (620-719) and high FICO (720+). Finally, Panel C plots the distribution of residuals from a regression of realized interest rates on borrower and loan characteristics. The black line residualizes against only borrower characteristics, which include the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. The light blue line plots residuals from a regression of rates on these borrower characteristics as well as lender $\times$ origination quarter fixed effects.

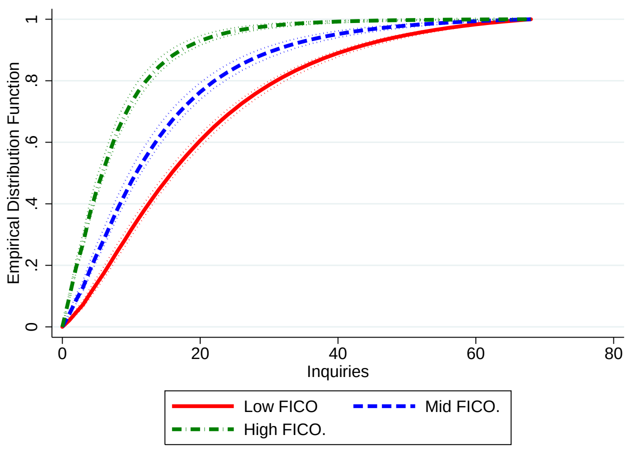
Figure 2: Inquiry distribution among mortgage applicants



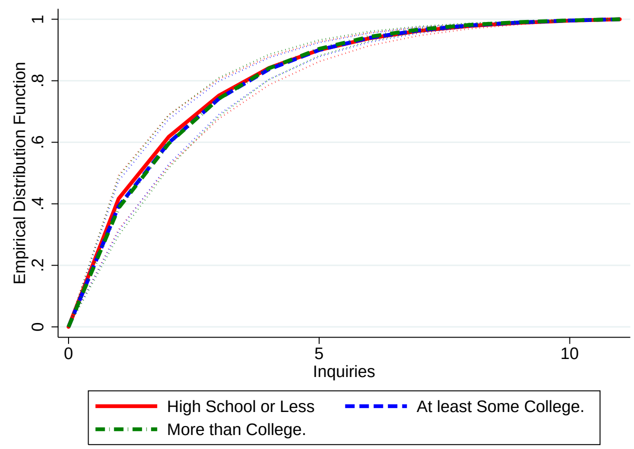
Panel A: Loan-Level Dataset



Panel B: Applicant Dataset



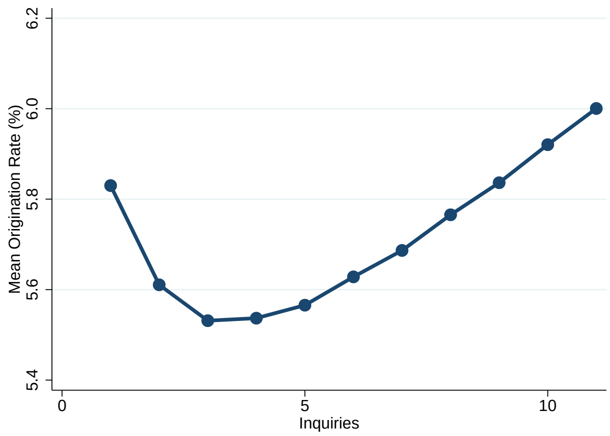
Panel C: Applicants by FICO Score



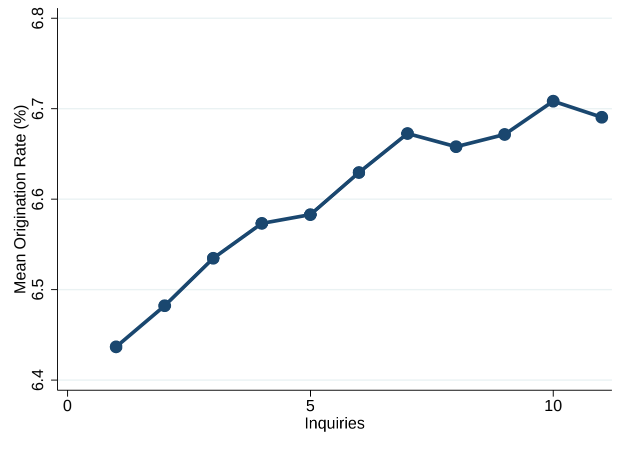
Panel D: Applicants by Education

Notes: Figure plots distribution of inquiries across successful mortgage applicants (i.e. those in our loan-level dataset) across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post crisis period from the first quarter of 2010 on. Dashed lines plot bootstrapped 95% confidence intervals, in which bootstraps are clustered at the origination quarter level. Panel A plots the inquiry distribution for all borrowers in our application-level dataset, while Panel B plots the inquiry distribution for our loan-level dataset containing borrowers with successful application. Panel C plots the distribution of inquiries across mortgage applicants for three FICO buckets: low FICO ($\leq$ 620), middle FICO (620-719) and high FICO (720+). Panel D plots the distribution of inquiries across successful mortgage applicants (i.e. those in our loan-level dataset) for three borrower education groups.

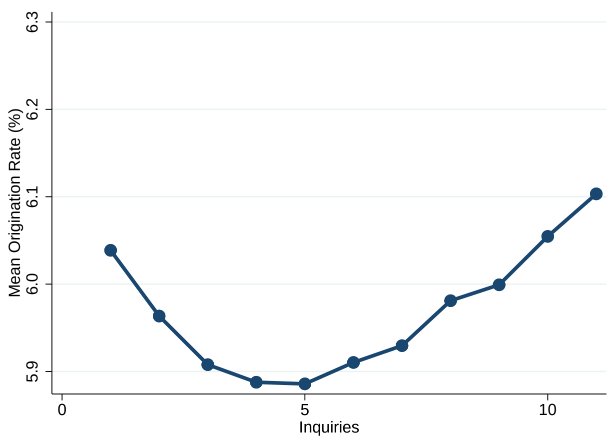
Figure 3: Rates and search by FICO bucket



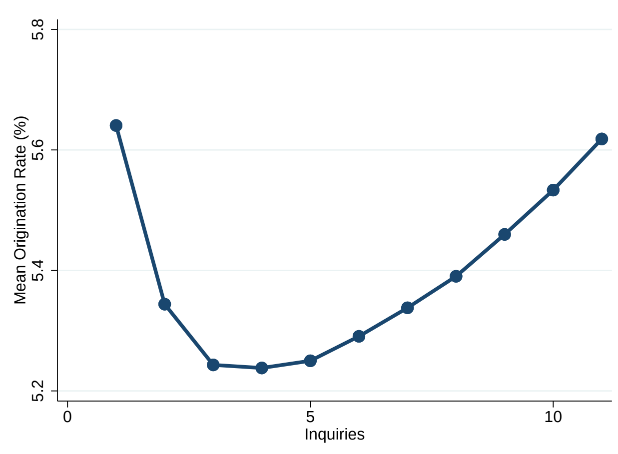
Panel A: All Borrowers



Panel B: FICO $\leq$620



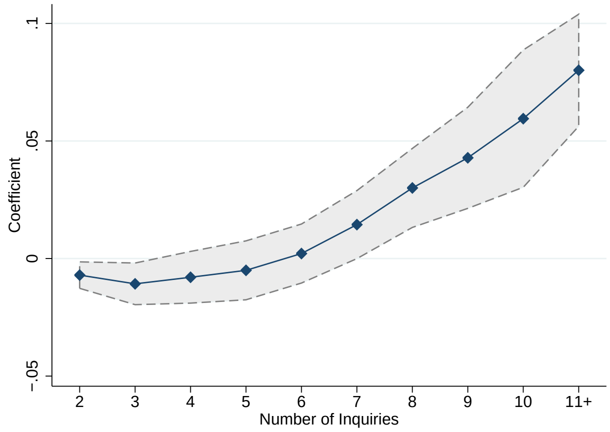
Panel C: 620 $<$ FICO $\leq$ 720



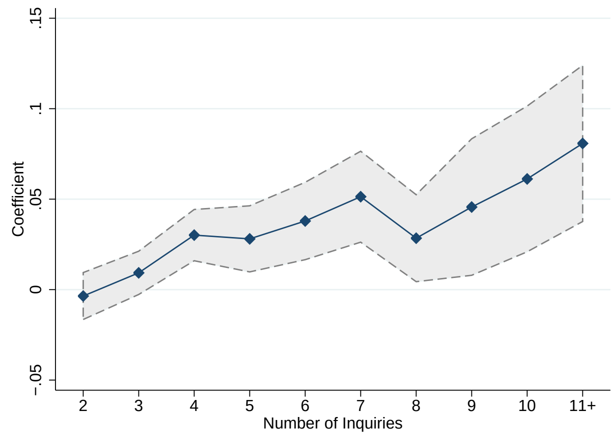
Panel D: FICO $>$ 720

Notes: Figure plots average realized interest rates against inquiry counts for realized loans for all borrowers and across three FICO buckets: low FICO ($\leq$ 620), middle FICO (620-719) and high FICO (720+).

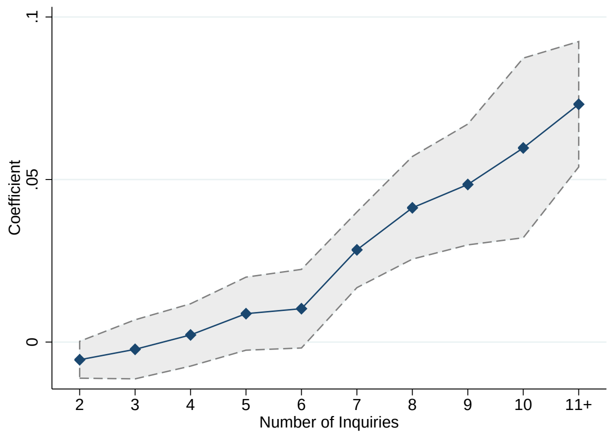
Figure 4: Relationship Between Search and Mortgage Origination Rates, Conditional on Observables



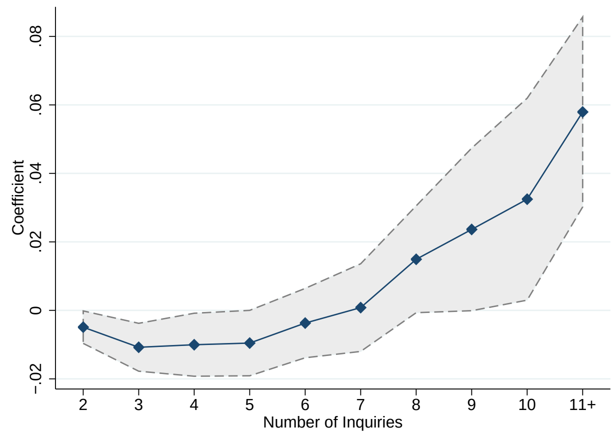
Panel A: All Borrowers



Panel B: FICO $\leq$620



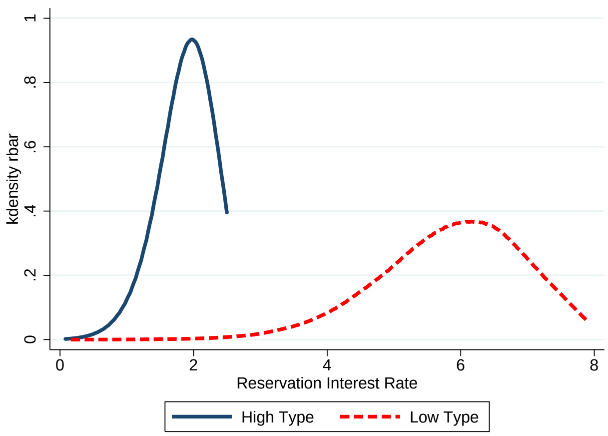
Panel C: 620 $<$ FICO $\leq$ 720



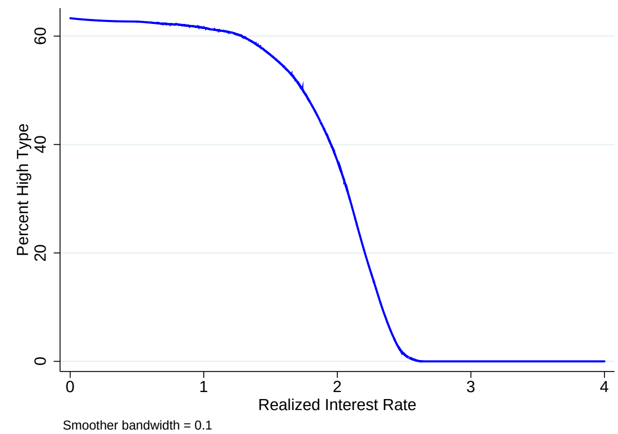
Panel D: FICO $>$ 720

Notes: Figure plots regression coefficients estimated from equation 2 using OLS across three FICO sub-samples. The dependent variable in each regression is the origination interest rate plus points and fees on a loan. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals. Panel A plots coefficients estimated from the full sample of borrowers, while Panels B, C, and D plot the coefficients estimated on the subsample of borrowers with FICO scores less than 620, between 620 and 720, and above 720, respectively.

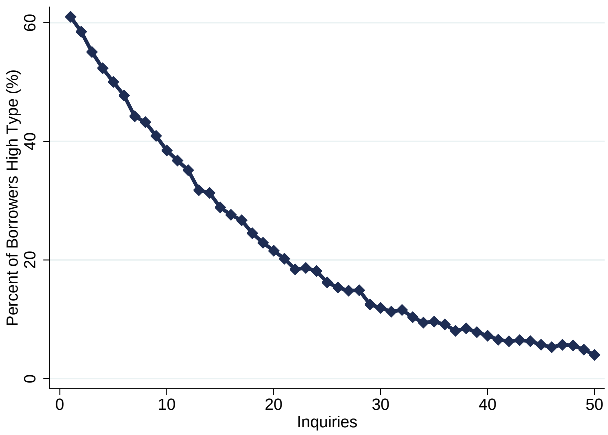
Figure 5: Characteristics of a Sequential Search Model with Informative Screening



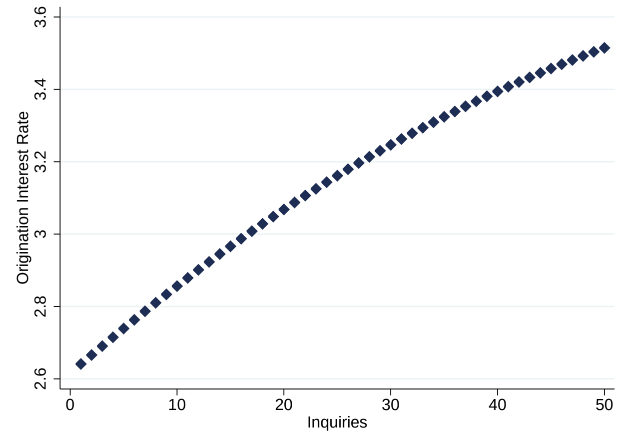
Panel A: Reservation rates of high and low types



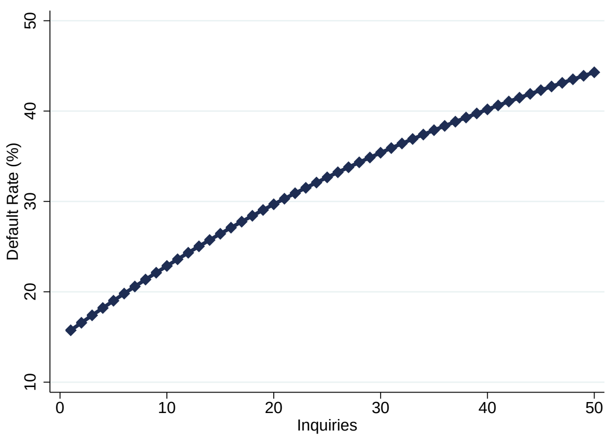
Panel B: Share of high types as function of origination rate



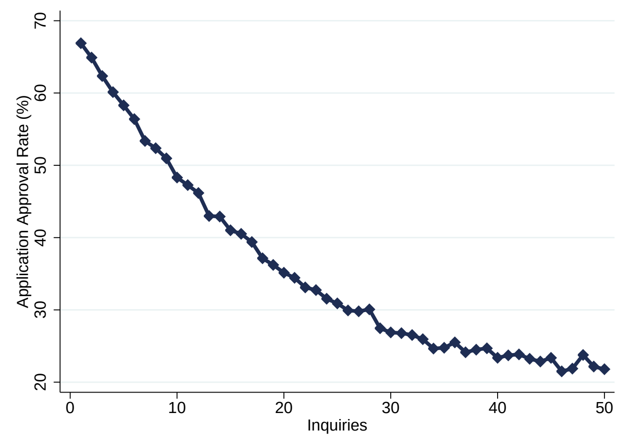
Panel C: Borrower type distribution and search



Panel D: Relationship between search and prices



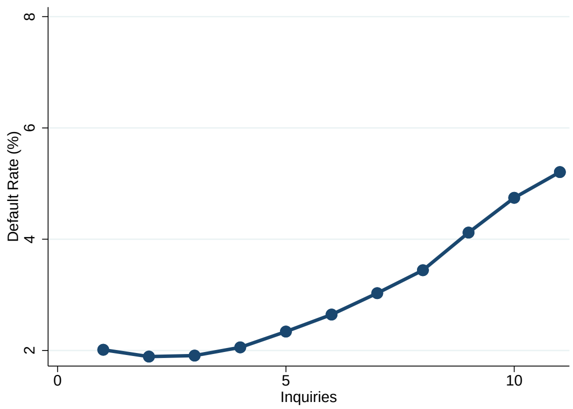
Panel E: Relationship between search and default rate



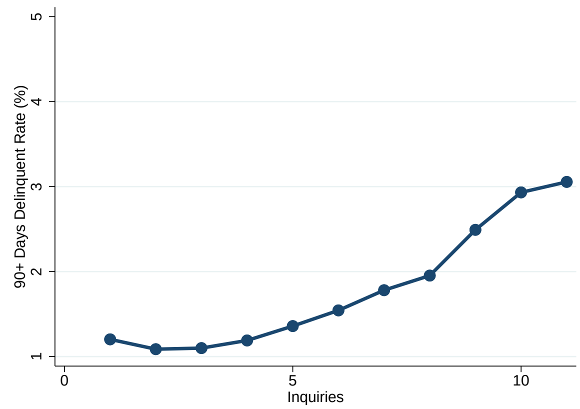
Panel F: Relationship between search and application approval

Notes: Figure plots key aspects of the mortgage market under the baseline model with informative screening. Data are simulated from a model in which application approval parameters are set to $p_h = 0.95$ and $p_l = 0.05$, the share of high types is $\lambda = 0.7$, the probability of full repayment for high and low types are $x_h = 0.8$, and $x_l = 0.4$, respectively, and the search costs and offered rates are distributed according to truncated normal distributions. Panel A plots the distribution of reservation rates for high type (in blue) and low type (in red) borrowers. Panel B plots the percent of borrowers that are high type at each realized interest rate, highlighting the pattern of adverse selection when screening is present. Panel C shows the percentage of successful borrowers who are high type as a function of search. Panel D, E, and F display the relationship between search and realized interest rates, eventual mortgage default rate, and application approval probability, respectively.

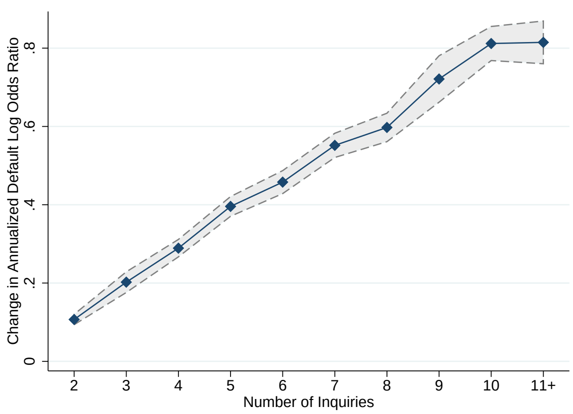
Figure 6: Search and Annualized Default Rate



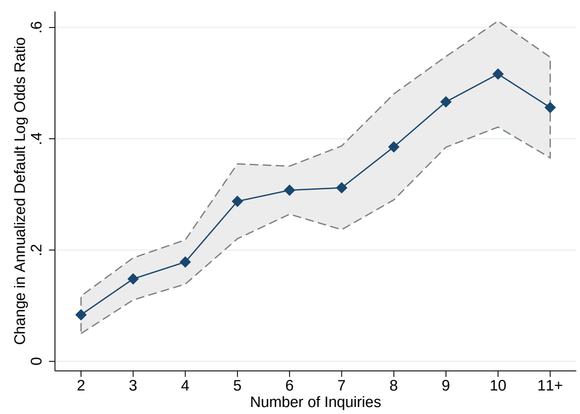
Panel A: Default



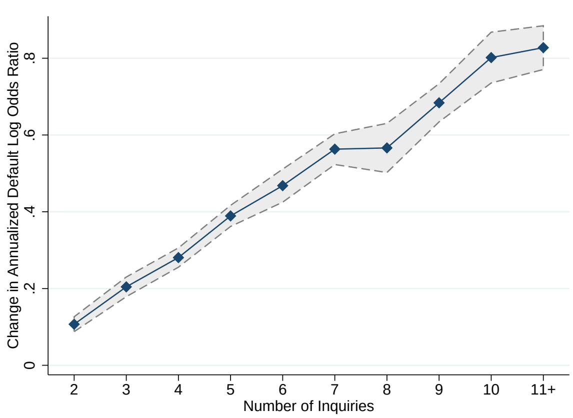
Panel B: 90+ Days Delinquent



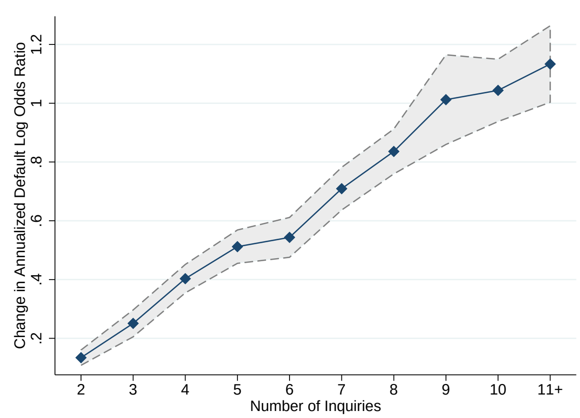
Panel C: All Borrowers, conditional on controls



Panel D: FICO $\leq$620, conditional on controls



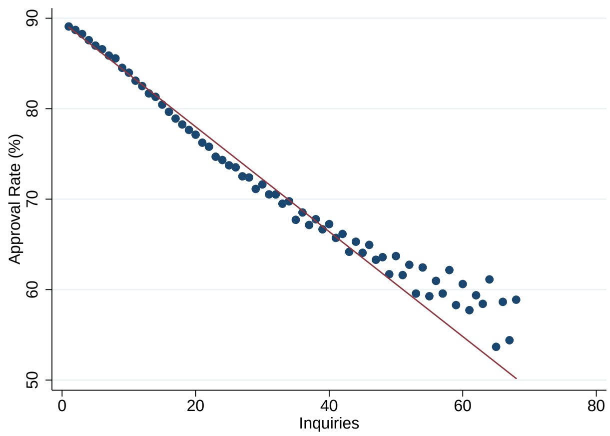
Panel E: 620 $<$ FICO $\leq$ 720, conditional on controls



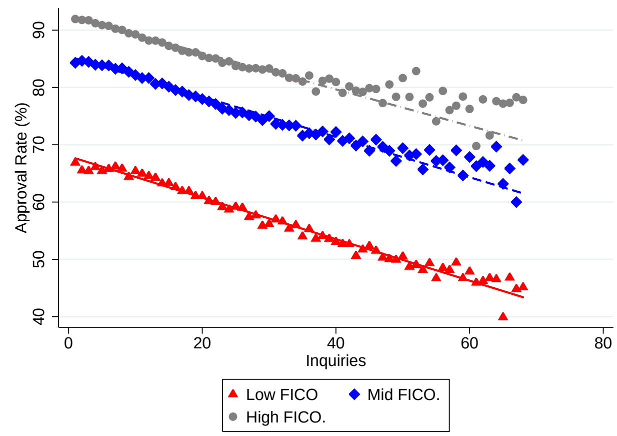
Panel F: FICO $>$ 720, conditional on controls

Notes: Figure plots average default rates against search. Panel A defines default to be serious (90+ days) delinquency, or foreclosure, while Panel B limits attention to seriously delinquent loans. Panels C through F plot regression coefficients estimated from equation 8 using MLE. The coefficients reflect changes in the log odds ratio of the annual default hazard relative to borrowers with one inquiry. Default is defined by the loan being at least 90 days delinquent, or entering foreclosure. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals. Panel C plots coefficients estimated from the full sample of borrowers, while Panels D, E, and F plot the coefficients estimated on the subsample of borrowers with FICO scores less than 620, between 620 and 720, and above 720, respectively.

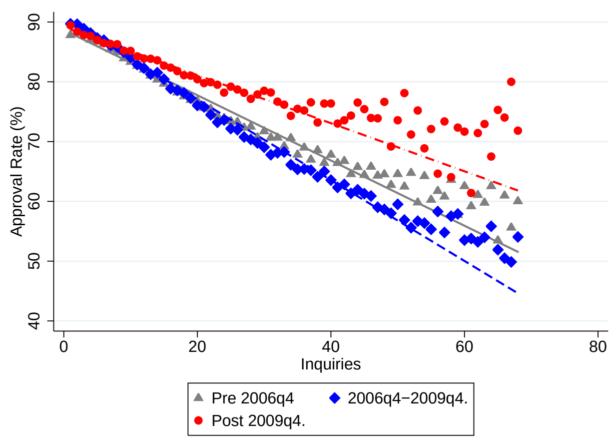
Figure 7: Relationship Between Search and Mortgage Application Approval Rates



Panel A: All Applicants



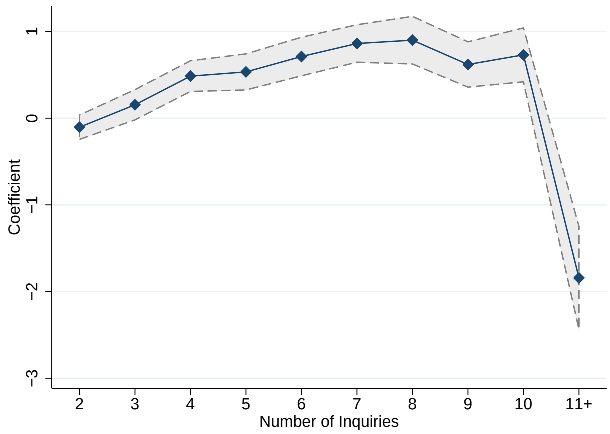
Panel B: By Borrower FICO Score



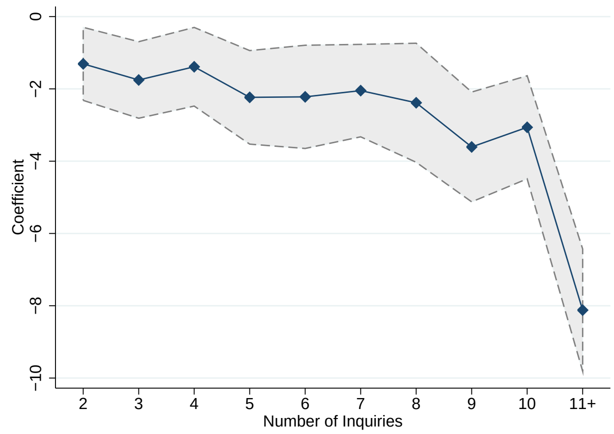
Panel C: By Origination Date

Notes: Figure plots the relationship between application approval rate and the number of inquiries on an applicant's credit report. A line of best fit, weighted by the number of applicants with s inquiries, is drawn as a visual aid. Panel A plots the relationship for all applicants in our application dataset. Panel B displays the relationship for three applicant FICO score buckets separately. The Low FICO group (in red) contains those with FICO score below 620, Mid FICO (in blue) corresponds to those with a FICO score between 620 and 720, while the High FICO group (in green) shows the patterns for those with a FICO score above 720. Panel C shows the patterns across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post crisis period from the first quarter of 2010 on. Lines of best fit, weighted by the number of applicants with s inquiries, drawn as a visual aid.

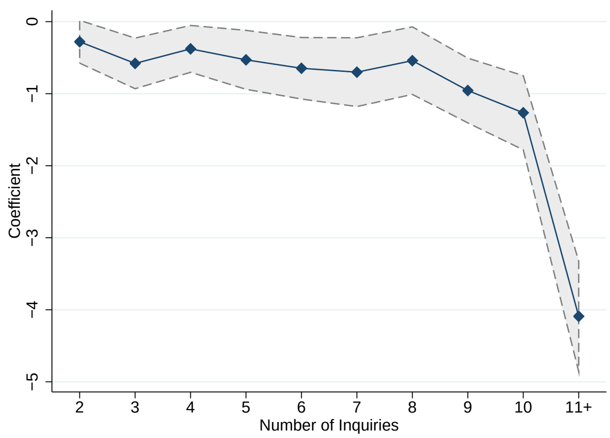
Figure 8: Relationship between search and mortgage application approval rates, conditional on observables by FICO bucket



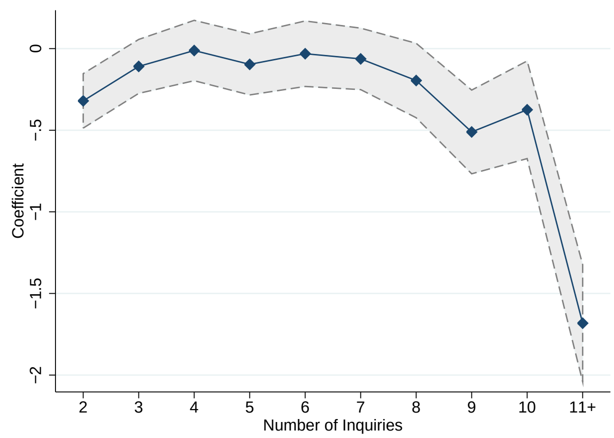
Panel A: All Applicants



Panel B: FICO $\leq$620



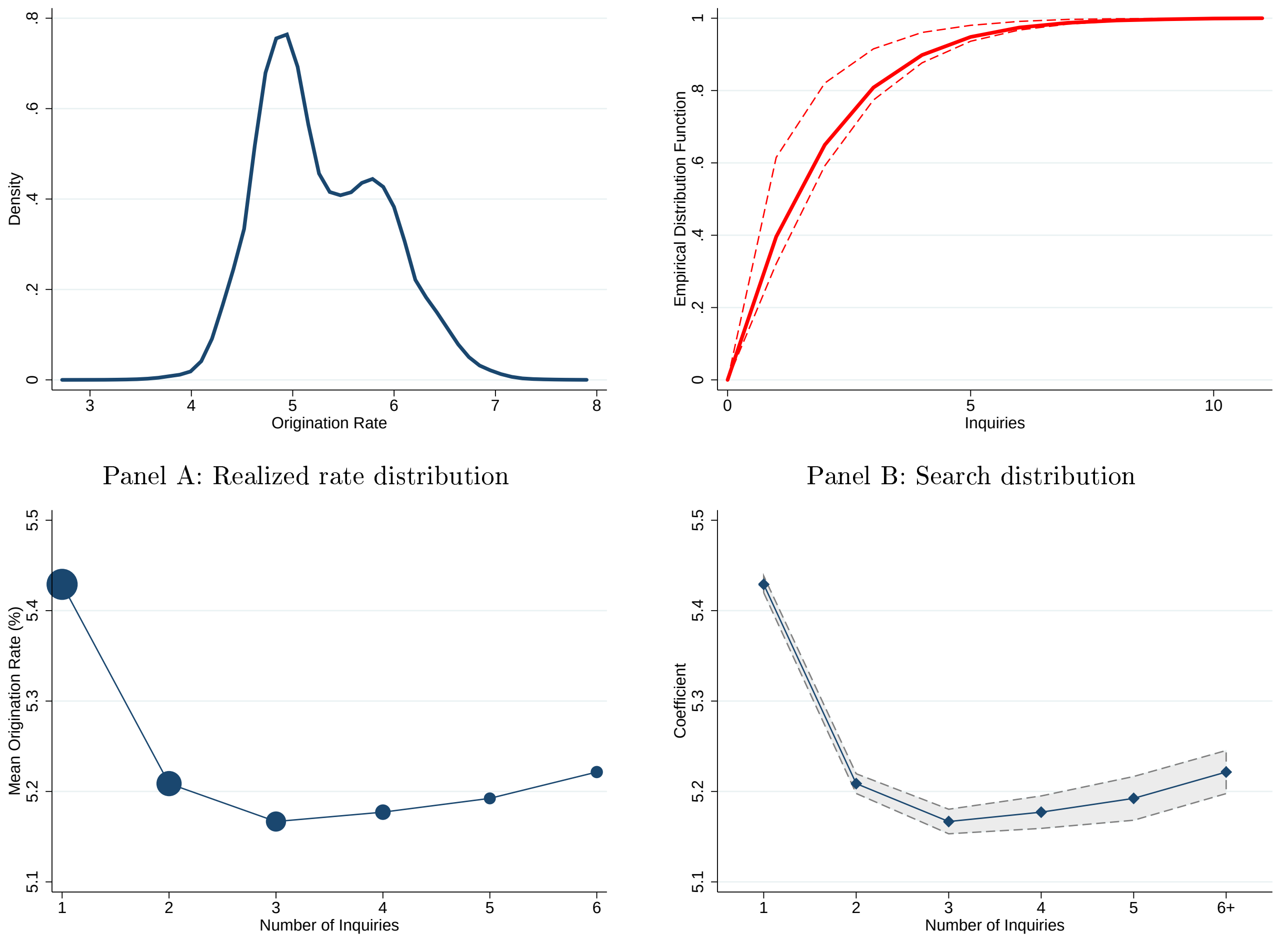
Panel C: 620 $<$ FICO $\leq$ 720



Panel D: FICO $>$ 720

Notes: Figure plots regression coefficients estimated from equation 9 using OLS. The dependent variable in each regression is an indicator for whether a mortgage application is approved, scaled by 100 for legibility. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals. Panel A plots coefficients estimated from the full sample of applicants, while Panels B, C, and D plot the coefficients estimated on the subsample of applicants with FICO scores less than 620, between 620 and 720, and above 720, respectively.

Figure 9: Search Behavior of Rarely Rejected Borrowers



Panel A: Realized rate distribution

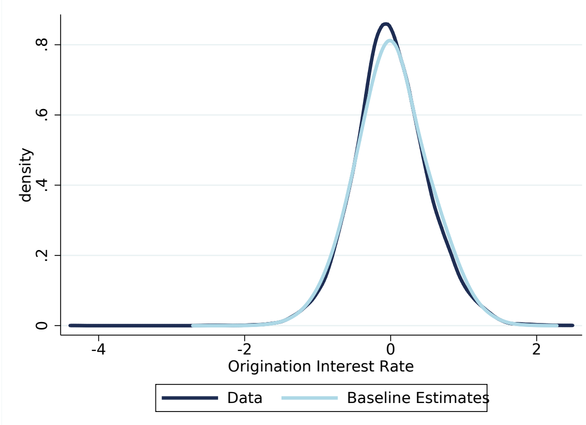
Panel B: Search distribution

Panel C: Relationship between search and prices

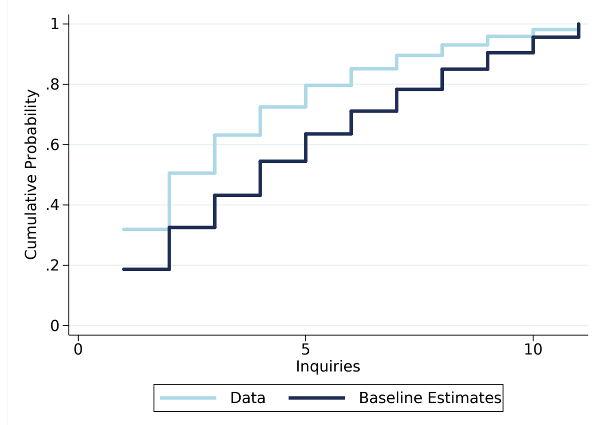
Panel D: Relationship between search and prices controlling for borrower observables

Notes: Figure plots key aspects of search behavior for a pool of borrowers whose applications are rarely rejected. Rarely-rejected borrowers are defined as those whose estimated propensity score from a logit regression on application approval status is above 0.975. All figures are produced using the dataset of realized loans. Panel A plots the estimated kernel density of realized interest rates for these borrowers. Panel B plots the distribution of inquiries for these borrowers. Dashed lines plot bootstrapped 95% confidence intervals, in which bootstraps are clustered at the origination quarter level. Panel C plots the mean origination interest rate as a function of the number of inquiries for this population of borrowers. The size of the marker for s inquiries is proportional to the number of rarely-rejected borrowers with s inquiries in the data. Panel D plots regression coefficients estimated from equation 2 using OLS, for a subsample of borrowers whose loan applications are rarely rejected. The dependent variable in theregression is the origination interest rate plus points and feeson a loan. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. White heteroskedasticity robust standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.

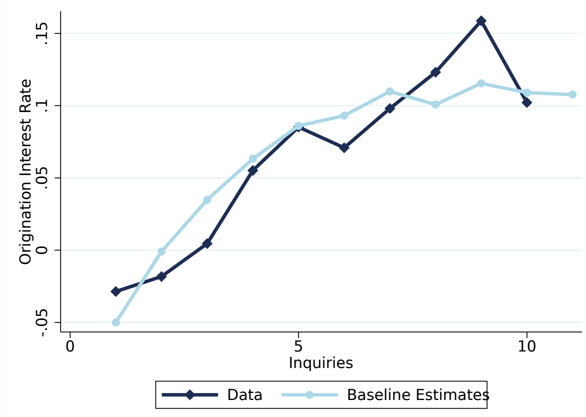
Figure 10: Model Performance: Search Behavior in Data Versus Model Simulation with Estimated Parameters



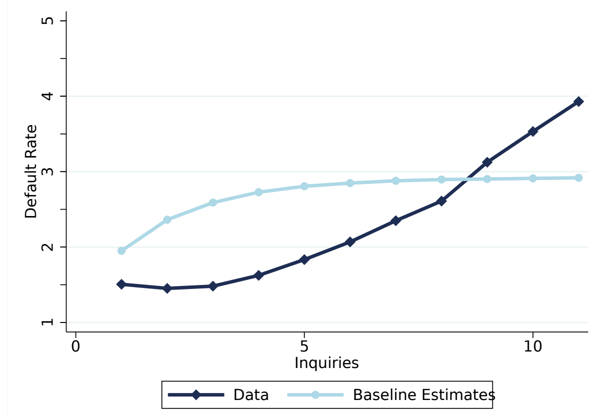
Panel A: Realized Residual Rate Densities



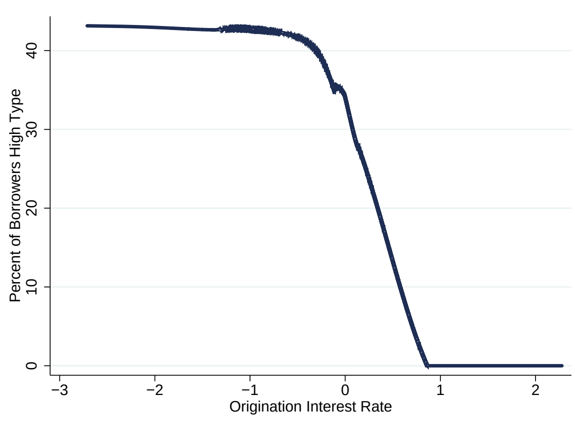
Panel B: Search Distribution



Panel C: Search and Origination Rates



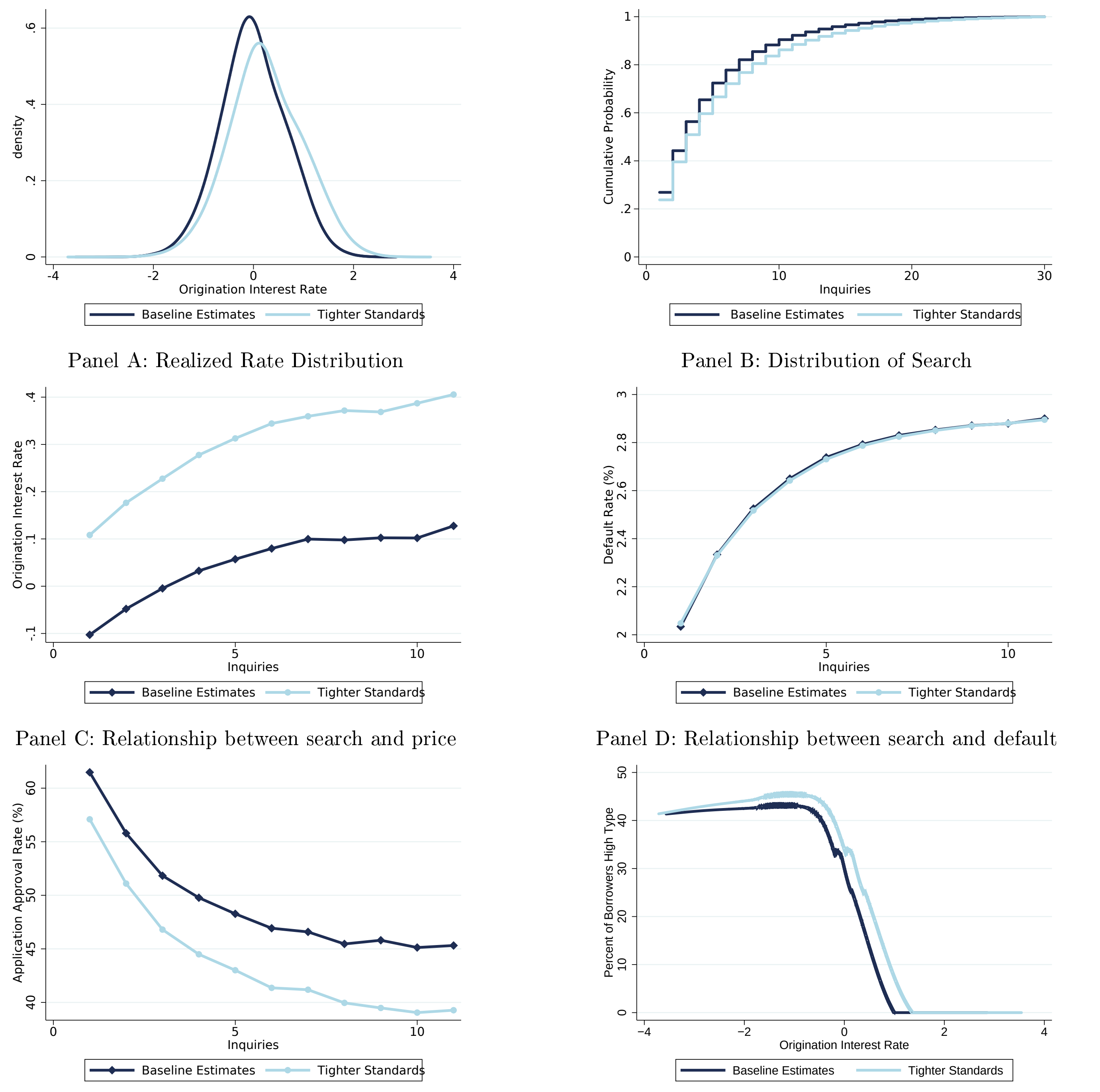
Panel D: Search and Default Probability



Panel E: Share of high types as function of origination rate

Notes: Figure plots the performance of our model under our benchmark estimated parameters from Table 5. Black lines plot quantities in our estimation sample, while light blue lines plot those implied by a large model simulation using parameters estimated by maximum likelihood following the approach laid out in section 7.1. Origination rates in data residualized against the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination year fixed effects. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C and D show the relationship between search and origination interest rates and default probability, respectively, where default probability is measured as of January 2015. To compute these default probabilities in the simulation, we randomly draw a mortgage's origination date from the distribution of origination dates in the data. Panel E shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate r who are of high type.

Figure 11: Tighter Lending Standards Counterfactual



Panel A: Realized Rate Distribution

Panel B: Distribution of Search

Panel C: Relationship between search and price

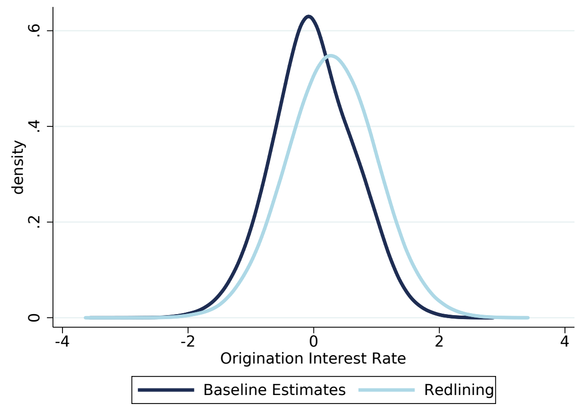
Panel D: Relationship between search and default

Panel E: Relationship between search and application approval

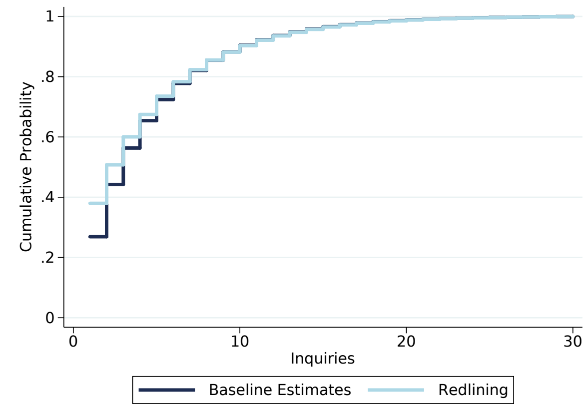
Panel F: Share of high types as function of origination rate

Notes: Figure plots the key aspects of search behavior under our baseline parameter estimates (black line) and a model in which the odds of application approval drop as they did following the recession (light blue line), allowing the equilibrium offered rate distribution to adjust. Odds of application acceptance presumed to decline by 21.8% for both high and low type borrowers.Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C, D, and E show the relationship between search and origination interest rates, the probability of ever defaulting, and the application approval rate. Panel F shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate r who are of high type.

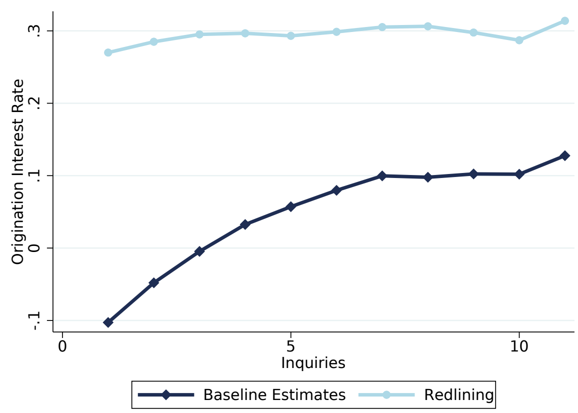
Figure 12: Redlining Counterfactual



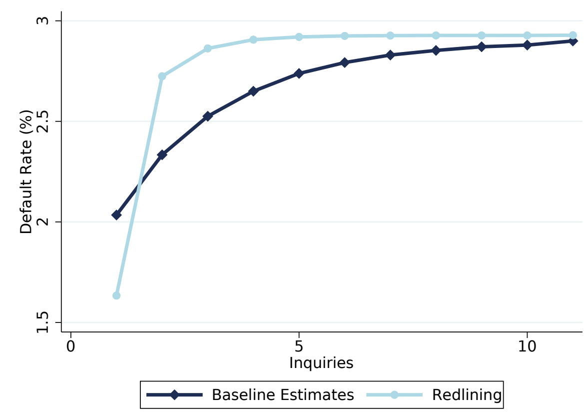
Panel A: Realized Rate Distribution



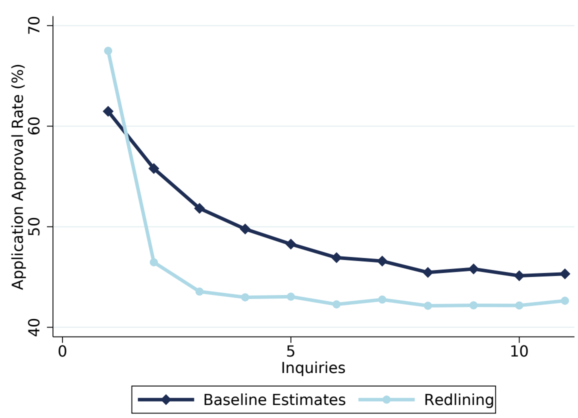
Panel B: Distribution of Search



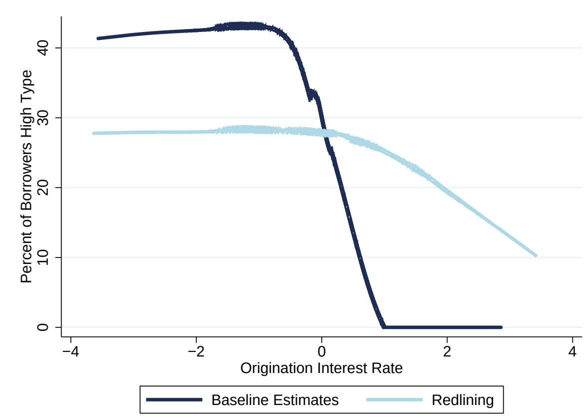
Panel C: Relationship between search and price



Panel D: Relationship between search and default



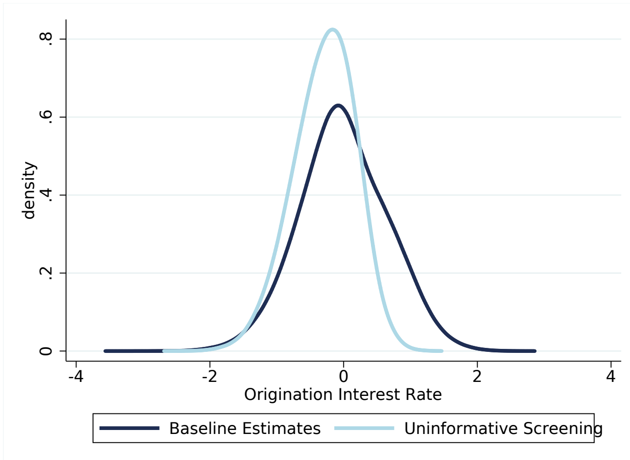
Panel E: Relationship between search and application approval



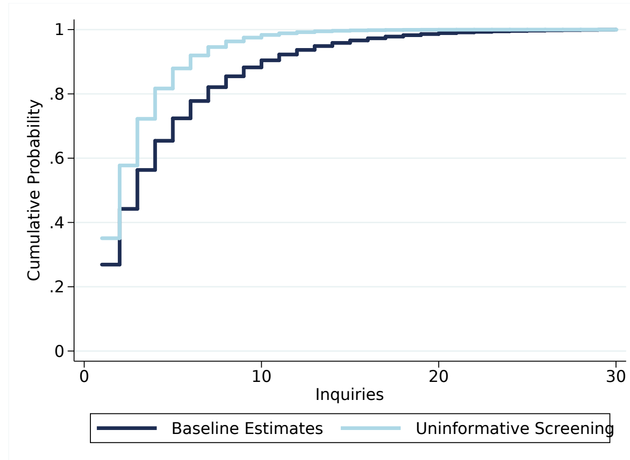
Panel F: Share of high types as function of origination rate

Notes: Figure plots the key aspects of search behavior under our baseline parameter estimates (black line) and the counterfactualmodel of redlining described in the text. We suppose that half of the lenders in the market engage in redlining behavior. Borrowers belong to two groups, B or W, which are identical in all respects, except redlining banks accept applications both high and low types in the B group at half of the rate that they accept applications from the W group. We let the B group comprise 20% of the pool of borrowers. Equilibrium offered rates are allowed to adjust. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C, D, and E show the relationship between search and origination interest rates, the probability of ever defaulting, and the application approval rate. Panel F shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate r who are of high type.

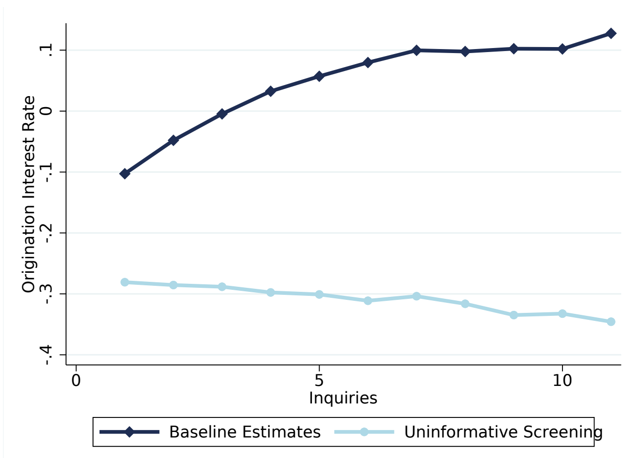
Figure 13: Place Based Policies Counterfactual



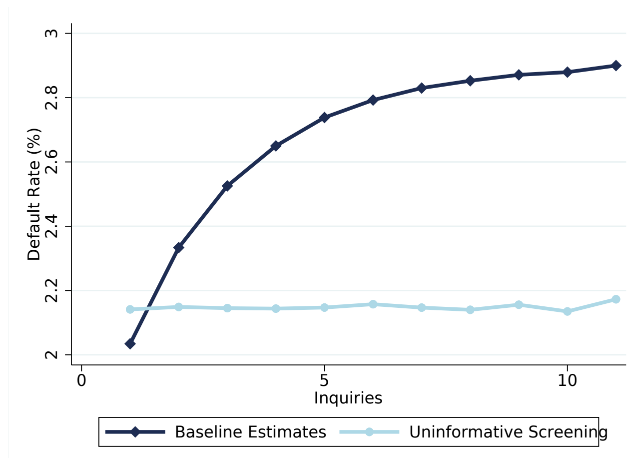
Panel A: Realized Rate Distribution



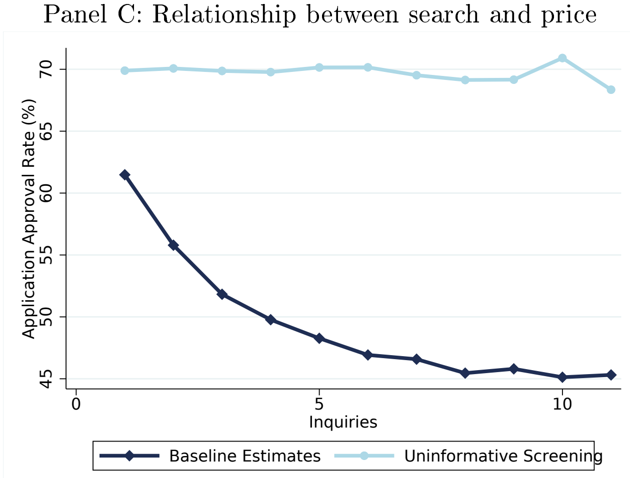
Panel B: Distribution of Search



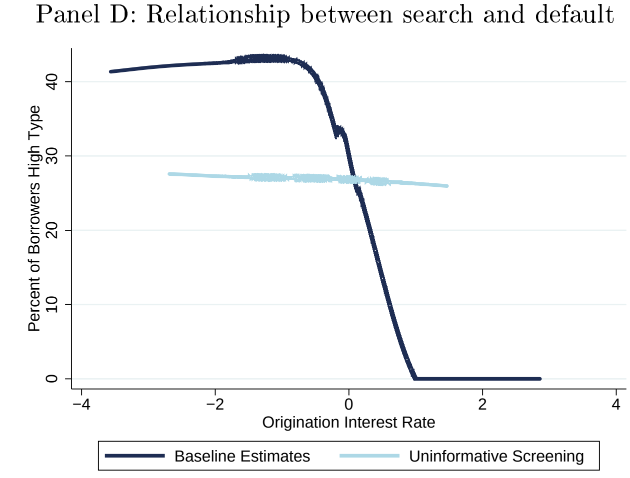
Panel C: Relationship between search and price



Panel D: Relationship between search and default



Panel E: Relationship between search and application approval



Panel F: Share of high types as function of origination rate

Notes: Figure plots the key aspects of search behavior under our baseline parameter estimates (black line) and a model in which applications from high and low type borrowers are rejected at the same rate, allowing the distribution of offered rates to adjust. This constant rate is given by the average approval probability under our baseline estimates: $\lambda p_h + (1-\lambda)p_l$. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C, D, and E show the relationship between search and origination interest rates, the probability of ever defaulting, and the application approval rate. Panel F shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate r who are of high type.

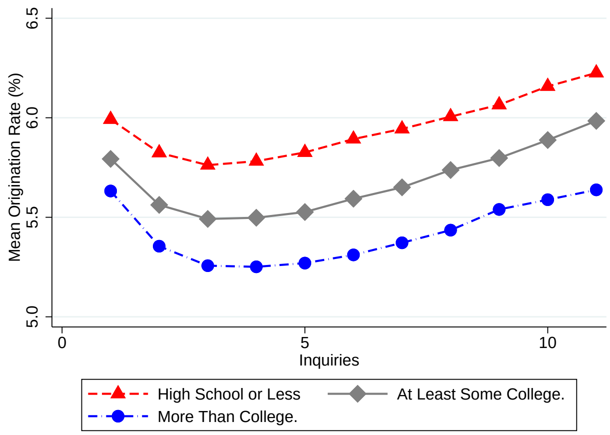
A Additional Robustness Tables and Figures

Table A1: k-means Clustering Test for Multiple Borrower Types

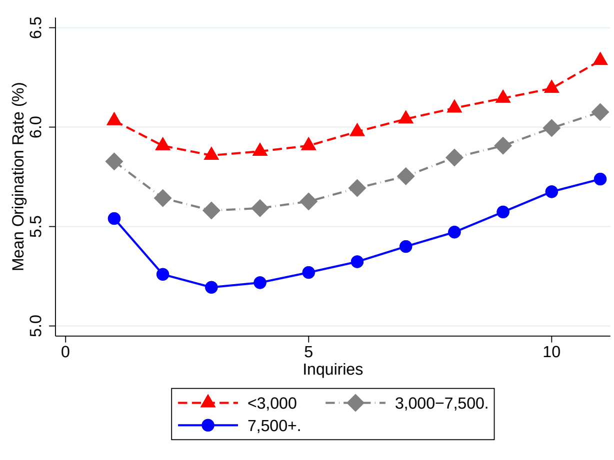
	OLS Residuals		Logit Residuals	
	(1)	(2)	(3)	(4)
	2 types	3 types	2 types	3 types
Default				
Type 1	0.000	0.000	0.000	0.000
	(1,106,792)	(474,524)	(1,106,792)	(406,650)
Type 2	1.000	0.000	1.000	0.000
	(210,015)	(632,268)	(210,015)	(700,142)
Type 3	-	1.000	-	1.000
	-	(210,015)	-	(406,650)
Approval				
Type 1	0.000	0.000	0.000	0.000
	(954,432)	(954,429)	(954,338)	(954,302)
Type 2	1.000	1.000	1.000	1.000
	(4,404,541)	(3,677,074)	(4,404,534)	(3,872,321)
Type 3	-	1.000	-	1.000
	-	(727,470)	-	(532,249)

Notes: Table shows the default and application approval probabilities within each k-means clustered group. Columns (1) and (3) impose that there are two latent types, while columns (2) and (4) assume three latent types. Columns (1) and (2) cluster individuals based on residuals from an OLS regression of an indicator for application approval or default on borrower observables, namely the borrower's FICO score, LTV ratio, back-end DTI ratio, product type, state and origination quarter fixed effects, refinance flags, and, for the default regressions, education, income and race. Columns (3) and (4) cluster individuals in a similar manner, only using a logit regression rather than OLS to estimate the probability of default or application approval. The size of each group is reported in parentheses beneath the default/approval rates.

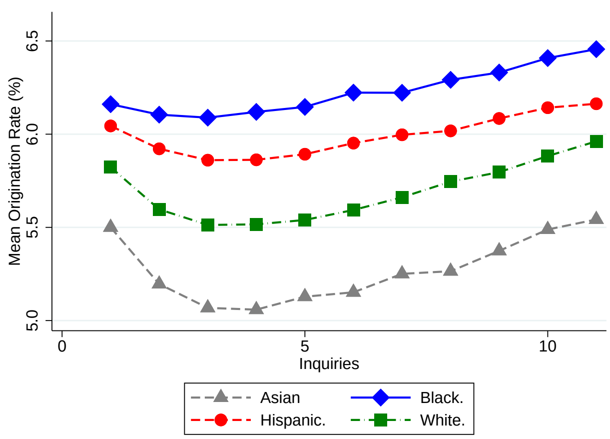
Figure A1: Mean Rates and Search by Borrower Observables



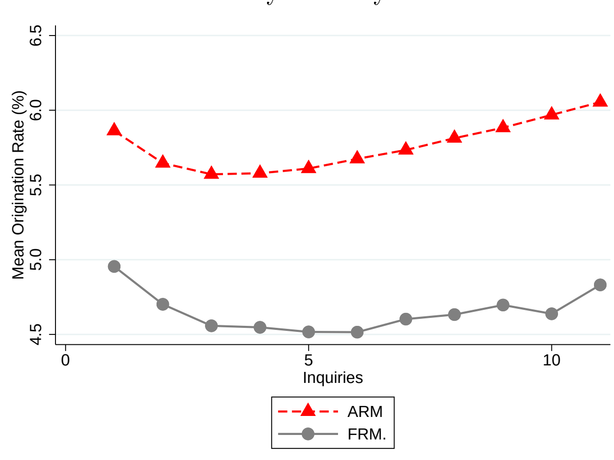
Panel A: By Education



Panel B: By Monthly Income



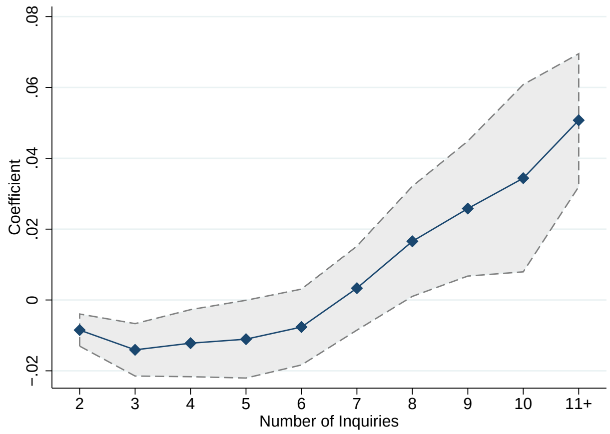
Panel C: By Race



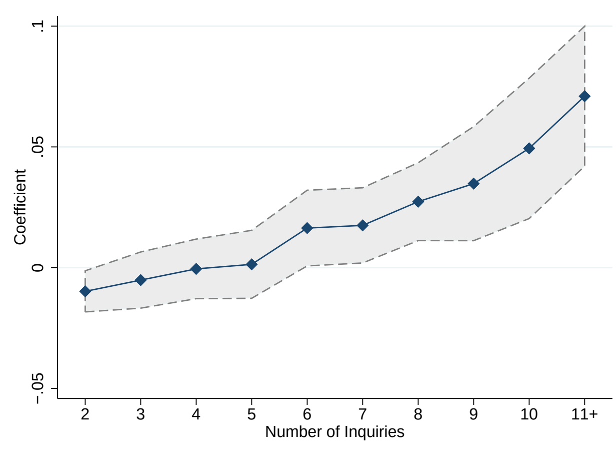
Panel D: By Product Type

Notes: Figure plots average realized interest rates against inquiry counts for realized loans across various borrower observables.

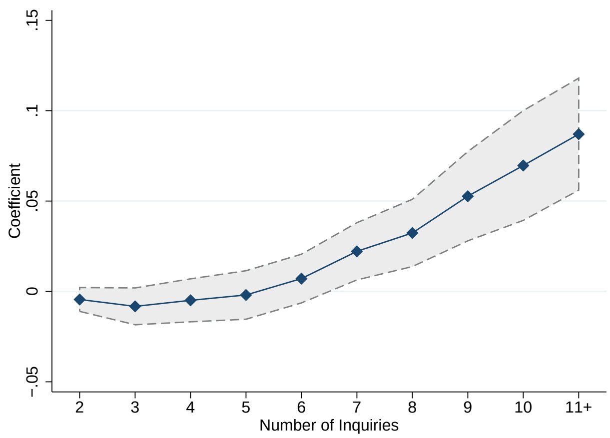
Figure A2: Relationship between Search and Realized Mortgage Interest Rates, Conditional on Observables, by Ex-Post Delinquency Status and Brokerage Status



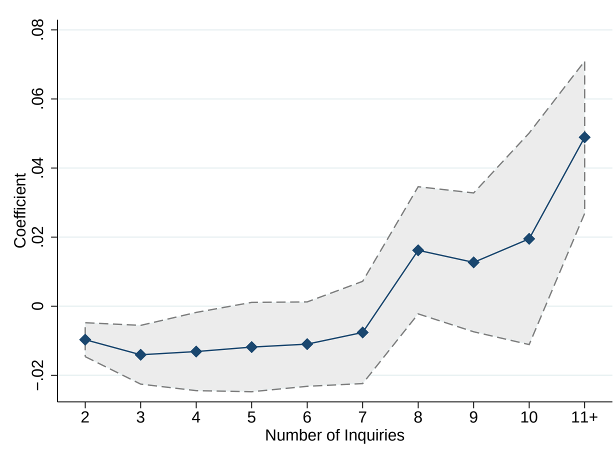
Panel A: Loans which do not default ex post



Panel B: Loans which default ex post



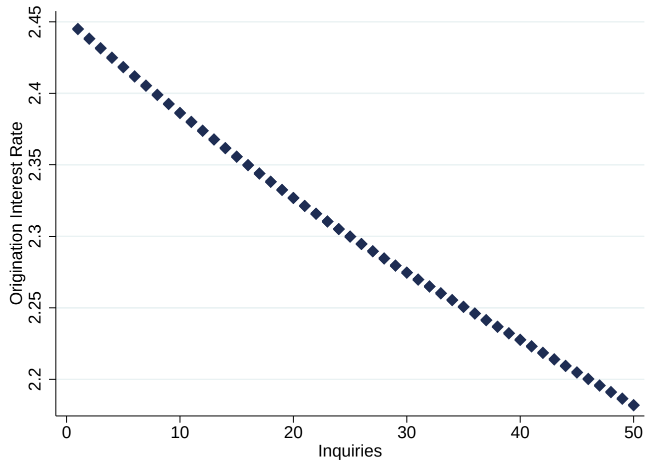
Panel C: Unbrokered loans



Panel D: Brokered loans

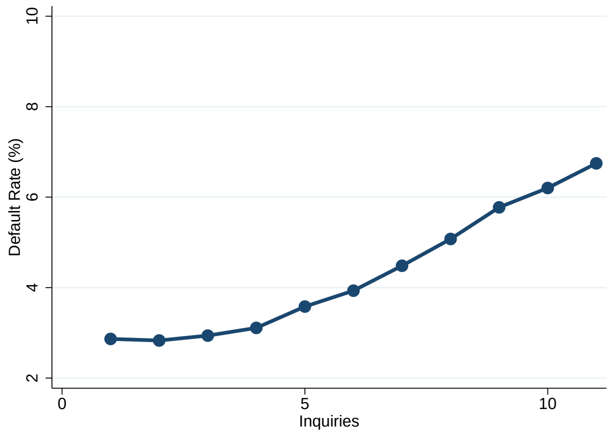
Notes: Figure plots regression coefficients estimated from equation 2 using OLS for the separate subsamples of loans which do not default ex post (Panel A), which do default ex post (Panel B), which are mortgages found without a broker (Panel C), and which are brokered mortgages (Panel D). The dependent variable in each regression is the origination interest rate on a loan. Default defined as a loan being in foreclosure or at least 90 days delinquent by Jan 1, 2015. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.

Figure A3: Relationship between search and prices in standard search model without screening

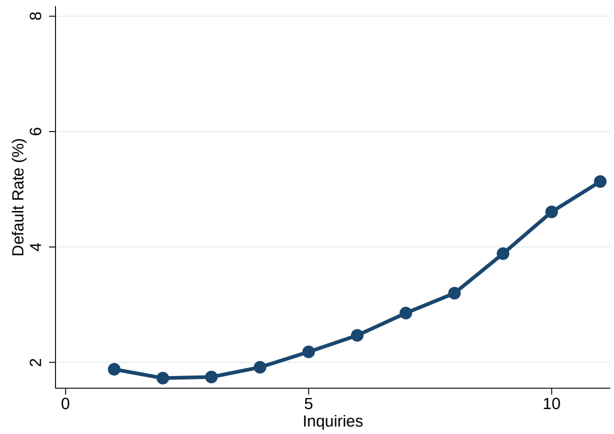


Notes: Figure plots the relationship between origination interest rates and search in the absence of screening: where $p_h = p_l = 1$, and the search costs and offered rates are distributed according to truncated normal distributions.

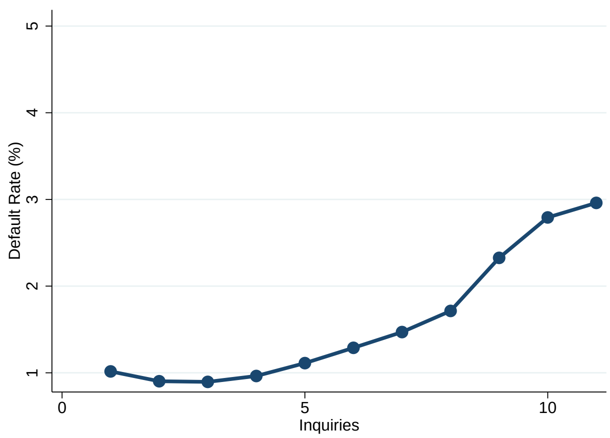
Figure A4: Default rate and search: by education and income level



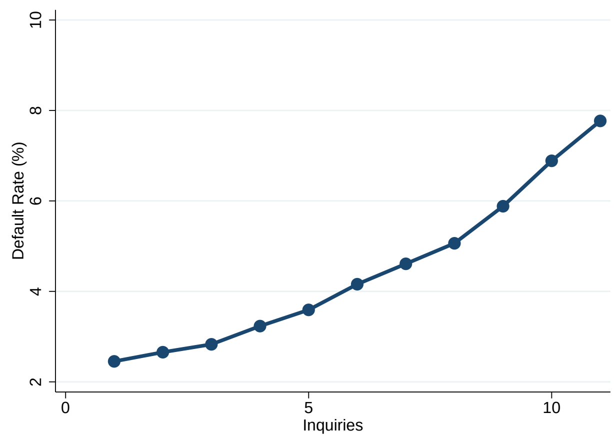
Panel A: High School or Less



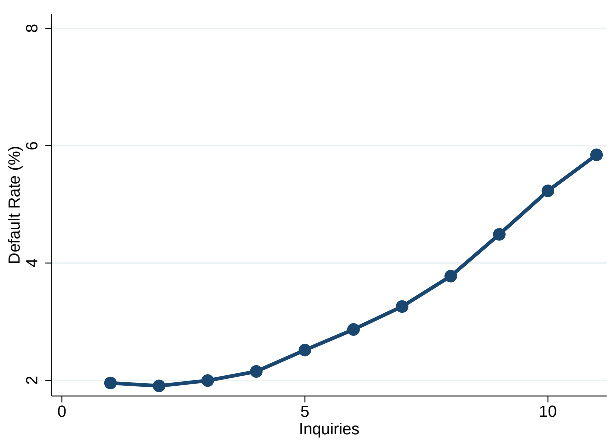
Panel B: Some College



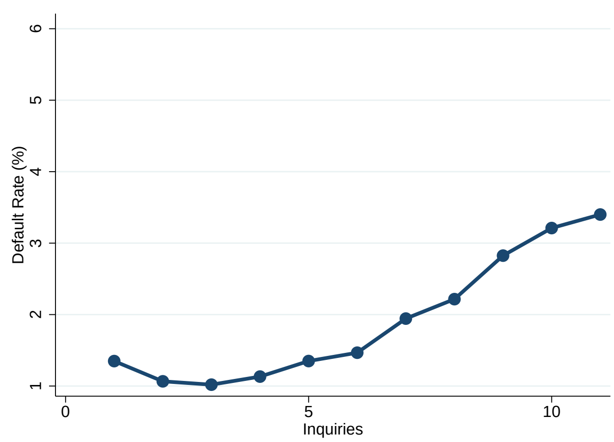
Panel C: College Graduate



Panel D: Monthly Income$\leq$\$3,000



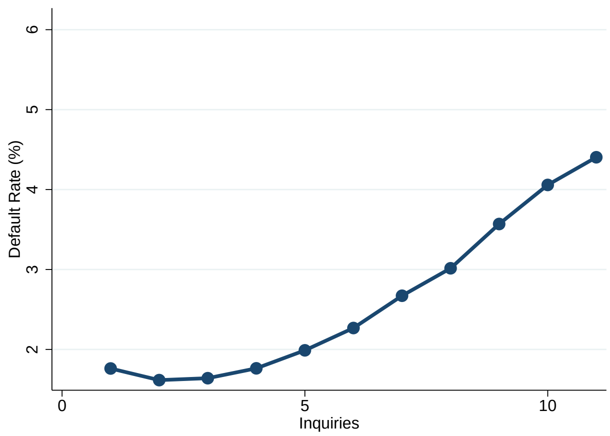
Panel E: \$3,000 $<$ Income$\leq$ \$7,500



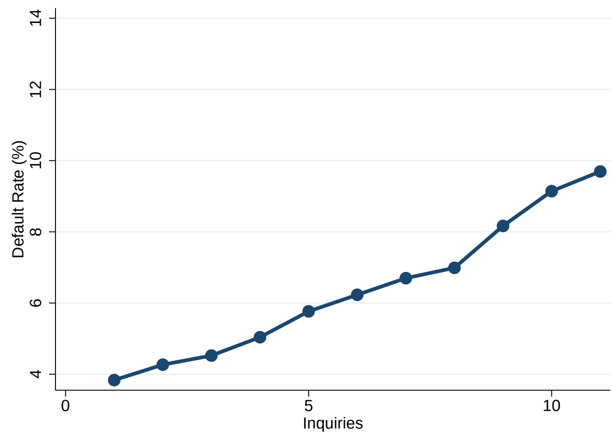
Panel F: Monthly Income $>$ \$7,500

Notes: Figures plots average annualized default rates against search, subsetting borrowers according to their education and monthly income levels.

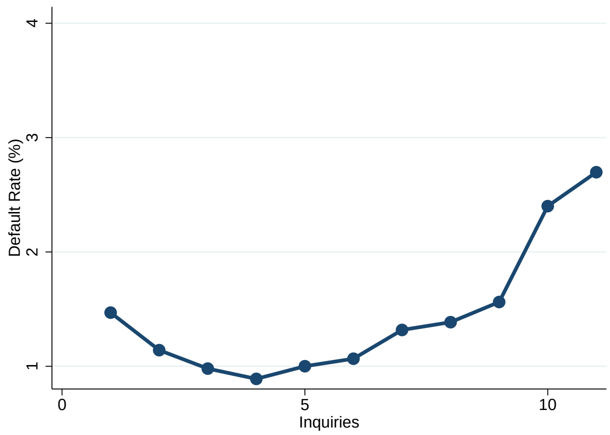
Figure A5: Default rate and search: by borrower race



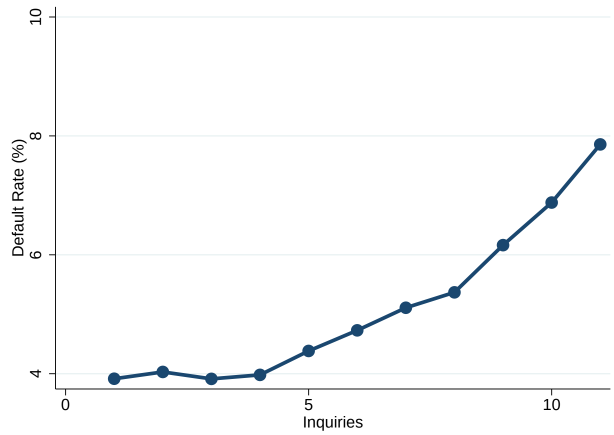
Panel A: White



Panel B: Black



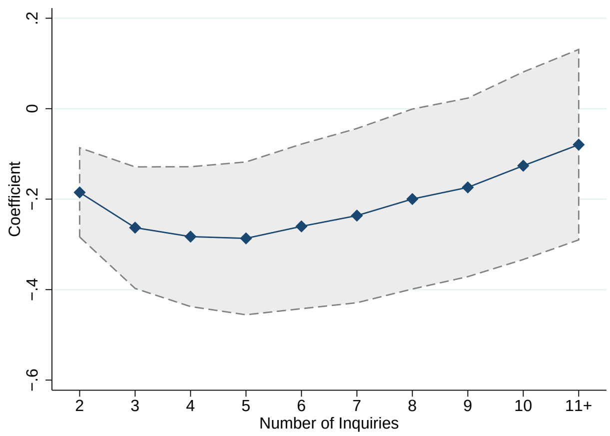
Panel C: Asian



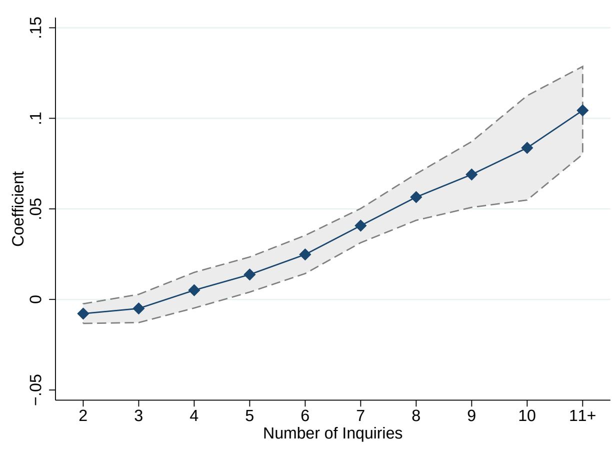
Panel D: Hispanic

Notes: Figures plots average annualized default rates against search, subsetting borrowers according to their race.

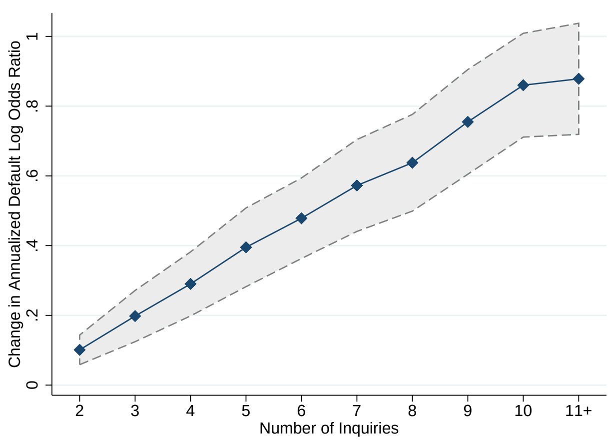
Figure A6: Relationship Between Search and Realized Interest or Default Rates, Controlling for LLPA Categories



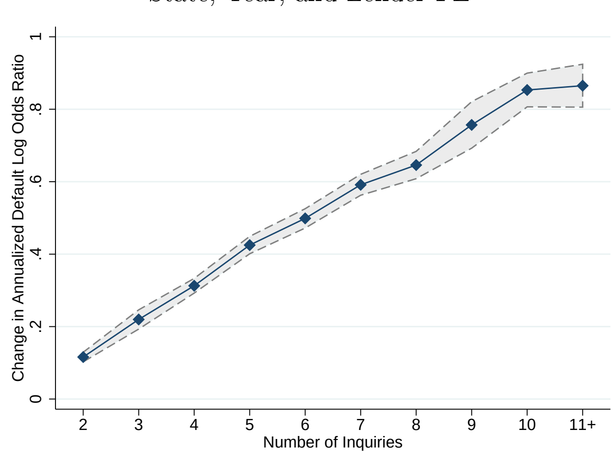
Panel A: Interest Rates
LLPA Categories Only



Panel B: Interest Rates
LLPA Categories +
State, Year, and Lender FE



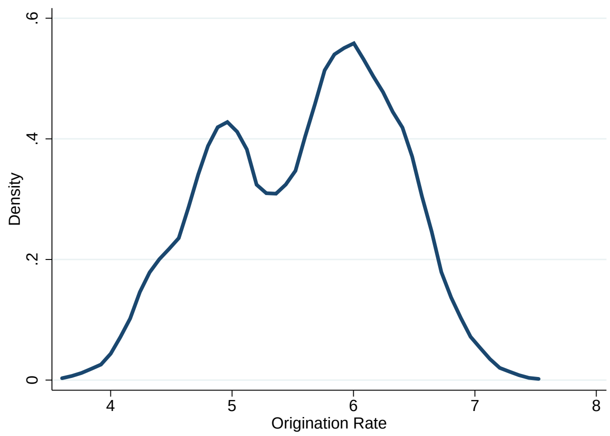
Panel C: Default Rates
LLPA Categories Only



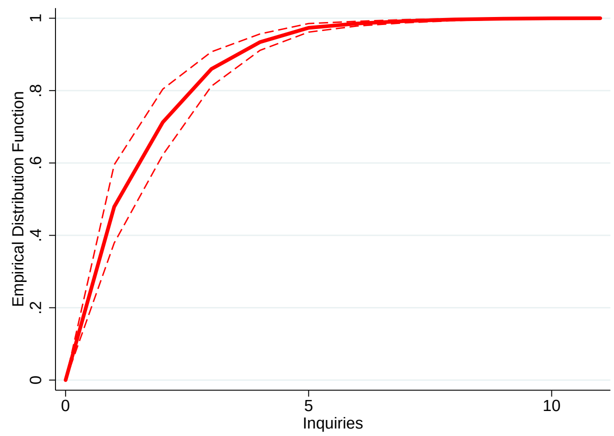
Panel D: Default Rates
LLPA Categories +
State, Year, and Lender FE

Notes: Panel A and B plots regression coefficients estimated from equation 2 using OLS. The dependent variable in Panels A and B is the origination interest rate on a loan. Panels C and D plot regression coefficients estimated from equation 8 using MLE. The coefficients reflect changes in the log odds ratio of the annual default hazard relative to borrowers with one inquiry. Default is defined by the loan being at least 90 days delinquent, or entering foreclosure. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. Controls are included for every bin for loan-level price adjustment as urged by Fannie Mae, available at `https://www.fanniemae.com/content/pricing/llpa-matrix.pdf`. Panels B and D additionally include state fixed effects, lender fixed effects, and origination quarter fixed effects. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.

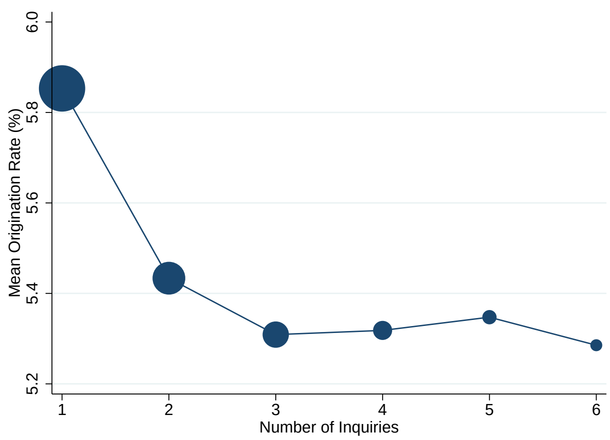
Figure A7: Search Behavior of Rarely Rejected Borrowers - Alternative Rarely Rejected Definition



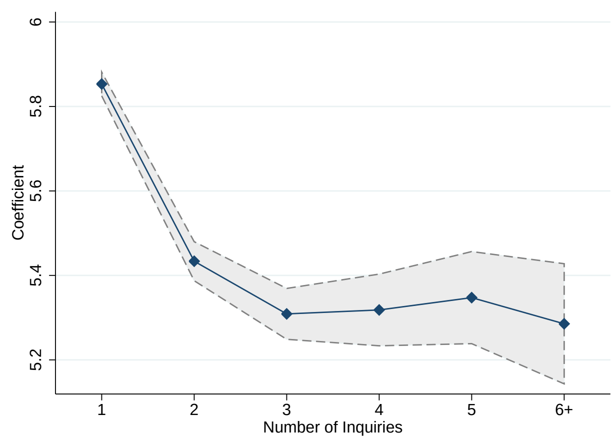
Panel A: Realized rate distribution



Panel B: Search distribution



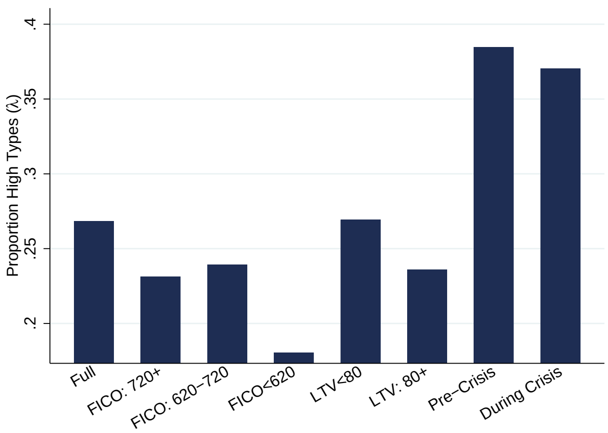
Panel C: Relationship between search and prices



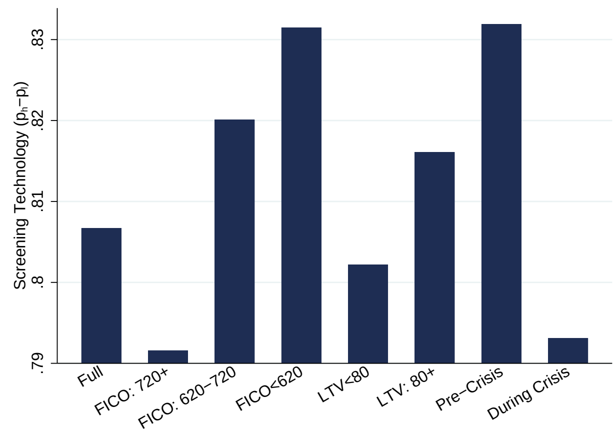
Panel D: Relationship between search and prices controlling for borrower observables

Notes: Figure plots key aspects of search behavior for a pool of borrowers whose applications are rarely rejected. Rarely-rejected borrowers are defined as thoseapplying for 30-year fixed rate mortgages with combined origination loan-to-value ratio below 60, DTI ratio below 40, FICO score above 800. All figures are produced using the dataset of realized loans. Panel A plots the estimated kernel density of realized interest rates for these borrowers. Dashed lines plot bootstrapped 95% confidence intervals, in which bootstraps are clustered at the origination quarter level. Panel C plots the mean origination interest rate as a function of the number of inquiries for this population of borrowers. The size of the marker for s inquiries is proportional to the number of rarely-rejected borrowers with s inquiries in the data. Panel B plots the distribution of inquiries for these borrowers. Panel D plots regression coefficients estimated from equation 2 using OLS, for a subsample of borrowers whose loan applications are rarely rejected. The dependent variable in each regression is the origination interest rate on a loan. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. White heteroskedasticity robust standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.

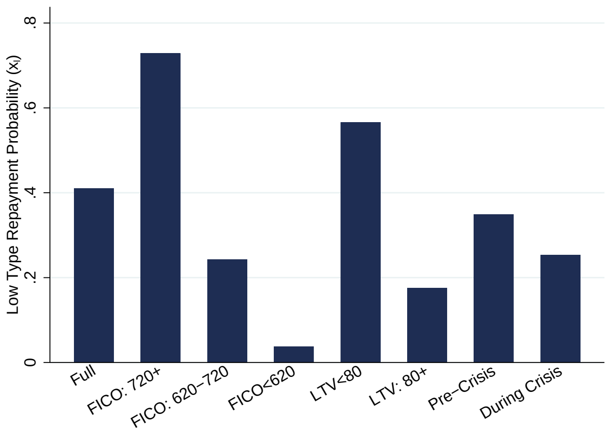
Figure A8: Estimates by subsample



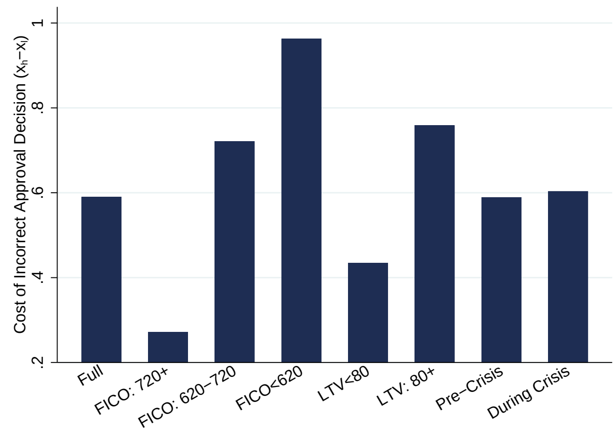
Panel A: Proportion of high types λ



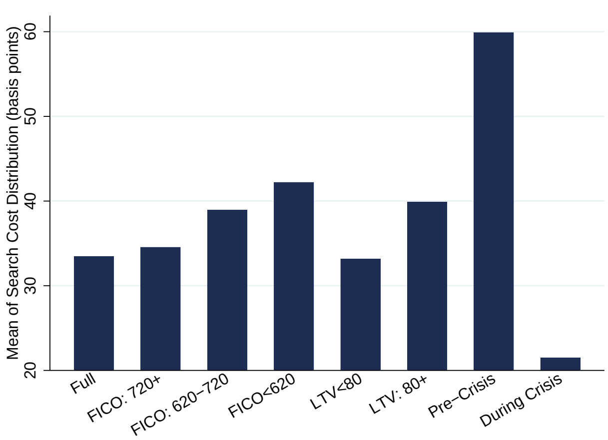
Panel B: Screening technology power $p_h - p_l$



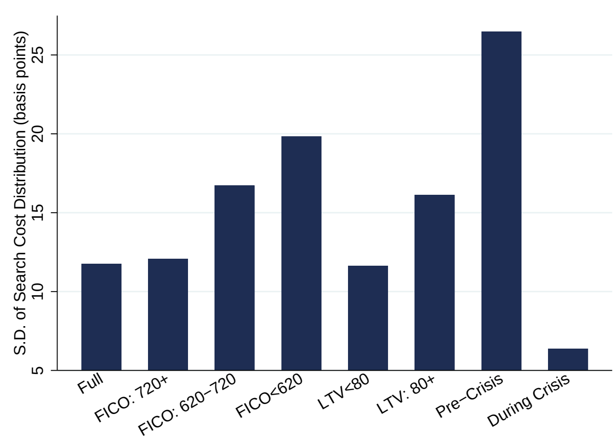
Panel C: Repayment probability of low types x_l



Panel D: Cost of misclassification $x_h - x_l$



Panel E: Mean search cost $e^{(\mu_c+\sigma_c^2/2)}$



Panel F: Standard deviation of search cost $\sqrt{\left(e^{\sigma_c^2}-1\right)e^{(2\mu_c+\sigma_c^2)}}$

Notes: Figure shows estimated parameter values from our maximum likelihood routine across 8 subsamples. The sample of borrowers originating their mortgage in 2010 or later is omitted due to small sample size. The acceptance probability for high types p_h is 1 for all subsamples.

B Testing Binary Type Assumption

Throughout our analysis, we have assumed that borrowers belong to one of two types: high types who repay their mortgage with high probability, and low types who are less likely to repay their loan. There is no a priori reason to suppose that borrowers can be classified in this simple binary manner. To test this assumption, we use insights developed in the machine learning literature. First, we regress an individual's probability of default and application approval on a vector of borrower, loan, and application characteristics, as well as state and time fixed effects, following equation 7. The residuals from these regressions may be interpreted as the unobservable (to the econometrician) determinants of default and application approval, analogous to the p_z and x_z of our model. With these residuals in hand, we employ a k-means clustering algorithm to group borrowers into two and three groups, respectively.

The results are presented in Table A1. The table presents, for each clustered group, the probability of default (panel A) and application approval (panel B) in the data. Columns 1 and 3 present the results for a binary grouping, while columns 2 and 4 allow for a trinary type space. We find no evidence for a trinary type space. In both trinary and binary groupings, we observe one group which always defaults and one which always pays off its loan. Similarly, there exists one group which is always approved for a loan, while another group is never approved.

One might be concerned that this is driven by our linear functional form assumption. Therefore, columns 3 and 4 present analogous results when we estimate an individual's probability of default or application approval using a logit regression. The similarity between trinary and binary groupings is robust to alternative function form assumptions.

C Robustness of Positive Relationship Between Search and Rates

Figures A1-A2 and A4-A6 present robustness of the key empirical fact of the paper, namely that realized interest and default rates increase in borrower search. Figure A1 plots the mean realized interest rate against against search for a host of borrower subsets - by race, education, income, and product type. Figure A2 plots the estimated regression coefficients from equation 2 for the subset of loans that do/do not eventually default, and for the set of borrowers who do/do not obtain their mortgage from a broker. In all cases, we find the positive relationship between search and interest rates. Figures A4 through A5 plot the mean default rates as a function of search by borrower race, and monthly income. Again, we consistently find a positive relationship between search and default. Finally, Figure A6 plots coefficients estimated from Equations 2 (Panels A and B) and 7 (Panels C and D), after increasing our set of controls to include every bucket for loan-level price adjustments provided by Fannie Mae.[42] Panels A and C show the results when we omit state, origination date, and lender fixed effects, while Panels B and D include our full suite of fixed effects. Without controlling for aggregate trends, the relationship between search and interest rates becomes noisier. However, controlling for borrower state and origination quarter recovers the positive relationship between interest rates, default, and search. This is unsurprising - the unobserved quality of borrowers is thought to have changed substantially over our sample period, and varies substantially across states. Overall, the central fact of the paper appear robust to all manner of control variables and across nearly all subsets of borrowers.

[42] These adjustment factors may be obtained from `https://www.fanniemae.com/content/pricing/llpa-matrix.pdf`.

D Likelihood Construction

D.1 Demand

In our model, an inquiry is a draw from the offered rate distribution. Let S_i denote a random variable equal to the number of inquiries on loan application i, and let A_{is} be an indicator for whether an application sent on the s^{th} search was accepted. Define R_i to be the realized rate on mortgage i, and R_i^* to be the borrower i's reservation rate. Let D_i be an indicator for whether borrower i defaults on the mortgage. Finally, we let the random variable O_{is} denote the mortgage rate offered to (not necessarily applied for or realized by) borrower i on inquiry s, which has CDF $H(o)$.

We proceed using a maximum likelihood approach. First consider the probability that a realized loan with s inquiries, and origination interest rate r is observed. For the loan to have been realized on the s^{th} inquiry, the borrower must have failed to originate a mortgage on her first $s-1$ inquiries, and then observed a loan offered at rate r, applied for it, and had her application approved. To build the likelihood for such a borrower, suppose first that one could observe both the borrower's underlying type z and reservation rate r^*. The probability that the borrower originates a loan at a rate below r on her s^{th} inquiry is (suppressing the loan index i for legibility):

$$\begin{aligned}
Pr\left\{R \leq r, S=s|z,r^*\right\} &= Pr\{(\text{Applied and accepted for } s^{th} \text{ draw with rate} \leq r) \cap \text{ (Did not originate previously)}|z,r^*\} \\
&= Pr\{(\text{Offered rate less than } r \text{ and } r^* \text{ and accepted}) \cap \text{ (Did not originate previously)}|z,r^*\} \\
&= Pr\left\{(O_s \leq r \leq r^* \cap A_s = 1) \cap \left[\neg(O_1 \leq r^* \cap A_1 = 1) \cap \ldots \cap \neg(O_{s-1} \leq r^* \cap A_{s-1} = 1)\right]|z,r^*\right\} \\
&= Pr\left\{(O_s \leq r \leq r^* \cap A_s = 1)|z,r^*\right\} \left(Pr\left\{\neg(O \leq r^* \cap A = 1)|z,r^*\right\}\right)^{s-1} \\
&= \mathbf{1}\left\{r \leq r^*\right\} \cdot p_z H(r) \left(1 - p_z H(r^*)\right)^{s-1}
\end{aligned}$$

where $\neg$ represents logical negation. The third equality acknowledges that the borrower may not originate a loan in a given inquiry either because the offered rate was too high or because her application were rejected. The fourth equality follows by the i.i.d. nature of both borrower quality signals and offered rate draws, which stems from the assumption of undirected search. The final equality relies on the independence of borrower signals and offered rate draws. One may take the derivative of the above expression with respect to r to derive a likelihood of realizing a loan at rate r after s inquiries, conditional on a borrower's type and reservation rate:

$$l\left(R=r, S=s|z,r^*\right) = \mathbf{1}\left\{r \leq r^*\right\} \cdot p_z h(r) \left(1 - p_z H(r^*)\right)^{s-1}$$

for $h(r)$ the probability density function (pdf) of the offered rate distribution evaluated at r. In reality, we do not observe the borrower's reservation interest rate r^* or type z. Thus to form a feasible likelihood, it is necessary to integrate over the borrowers' possible reservation rates and type. Letting χ_{is} be an indicator for whether borrower i applied for the loan offered to her on her s^{th} search, this yields the likelihood function for the joint distribution of origination rates and search:

$$
\begin{aligned}
l\left(R_i = r, S_i = s | A_{is} = 1, \chi_{is} = 1\right) &= \lambda p_h h(r) \int_r^\infty \left(1 - p_h H(r^*)\right)^{s-1} dF_h(r^*) \\
&+ (1-\lambda) p_l h(r) \int_r^\infty \left(1 - p_l H(r^*)\right)^{s-1} dF_l(r^*)
\end{aligned}
$$

for $F_z(r^*)$ the equilibrium distribution of reservation rates for a borrower of type z.

Observe at this stage that our likelihood function does not incorporate the observed information on borrower default. In the model, the probability that a type z borrower does not default throughout the life of the loan is x_z. In the data, however, we do not observe whether the borrower will default at any point; instead, we observe the borrower's payment status as of January 1, 2015. We therefore must convert the default probability observed in the data, D_i, to match the default concept employed in our model. To do so, we assume that defaults occur with a constant hazard. Specifically, we let the term of the loan be given by T, and the number of months since origination be given by t. For instance, a 30-year fixed rate mortgage originated in January 2014 would have $T = 30 \times 12 = 360$ and $t = 12$ in January 2015. We may then define the survival function of the loan to be

$$
\Omega(t|z, T) = x_z^{t/T}
$$

Observe that $\Omega(0|z, T) = 1$, and $\Omega(T|z, T) = x_z$ as desired. Since the default indicator D_i is assumed to be independent from search and acceptance decisions, conditional on borrower type, including this information into our likelihood function is straightforward. Let $d \in \{0, 1\}$ be a realization of the random variable D_i. A borrower of type z, who has seen a share t/T of his loan term elapsed by January 2015, realizes $D_i = 0$ with probability $x_z^{t/T}$, and $D_i = 1$ with probability $1 - x_z^{t/T}$. Thus we may write the likelihood of the joint distribution of our loan data $(S_i, R_i, D_i | A_{is} = 1, \chi_{is} = 1; t, T)$ as follows:

$$
\begin{aligned}
l(R_i = r, S_i = s, D_i = d | A_{is} = 1, \chi_{is} = 1, t, T) = \lambda \left(d(1 - x_h^{t/T}) + (1-d) x_h^{t/T} \right) p_h h(r) \int_r^\infty \left(1 - p_h H(r^*)\right)^{s-1} dF_h(r^*) \\
+ \underbrace{(1-\lambda)}_{Pr\{z=l\}} \underbrace{\left(d(1 - x_l^{t/T}) + (1-d) x_l^{t/T} \right)}_{Pr\{D_i = d | z = l; t, T\}} \underbrace{p_l h(r) \int_r^\infty \left(1 - p_l H(r^*)\right)^{s-1} dF_l(r^*)}_{Pr\{R_s = r, A_s = 1, s-1 \text{ Failed Searchs} | z = l\}} \qquad (10)
\end{aligned}
$$

In our application-level dataset, we may not incorporate information on offered rates or default into our likelihood function. Instead, we simply match the probability of a borrower having s inquiries given that she applied for the loan: $Pr\{S_i = s | \chi_s = 1\}$. Again, we can write this as the probability of having $s - 1$ failed inquiries, conditional on applying for the offered rate on the s^{th} inquiry. The conditional probability formula implies that this probability may be expressed as

$$Pr\{s-1 \text{ failed inquiries}|\chi_{is}=1\} = \frac{Pr\{s-1 \text{ failed inquiries} \cap \chi_{is}=1\}}{Pr\{\chi_{is}=1\}}$$

It is straightforward to show, following a similar argument to that above,that the numerator may be written as

$$\begin{aligned} Pr\{s-1 \text{ failed inquiries} \cap \chi_{is}=1\} &= \lambda \int H(r^*)\left(1-p_h H(r^*)\right)^{s-1} dF_h(r^*) \\ &+ (1-\lambda)\int H(r^*)\left(1-p_l H(r^*)\right)^{s-1} dF_l(r^*) \end{aligned} \tag{11}$$

That is, the probability of applying for the s^{th} inquiry is the probability that the s^{th} inquiry yields an offered rate that is less than the borrower's reservation rate, multiplied by the probability that the borrower did not both apply for a loan and have her application accepted on any of the previous $s-1$ draws from the rate distribution. Integrating over the borrower's reservation rate and possible type yields the above expression.

It remains to derive $Pr\{\chi_{is}=1\}$, which is the probability that the s^{th} inquiry enters our application data through a borrower application. First, suppose that one could observe a maximum of $\tilde{S}$ inquiries for any individual borrower, and that each inquiry is, ex ante, equally likely to be observed. Since we only observe applicants who have are yet to originate a mortgage, the probability that we observe inquiry s' is then

$$\begin{aligned} \frac{1}{\tilde{S}} Pr\{s'-1 \text{ failed inquiries} \cap \chi_{is'}=1\} &= \frac{1}{\tilde{S}}\lambda \int H(r^*)\left(1-p_h H(r^*)\right)^{s'-1} dF_h(r^*) \\ &+ \frac{1}{\tilde{S}}(1-\lambda)\int H(r^*)\left(1-p_l H(r^*)\right)^{s'-1} dF_l(r^*) \end{aligned}$$

We could have observed any of the borrower's inquiries up to $\tilde{S}$. The probability that we observe exactly the s^{th} inquiry in an application is thus the probability of observing the s^{th} inquiry, divided by the total probability of observing any inquiry up to $\tilde{S}$:

$$Pr\{s-1 \text{ failed inquiries}|\chi_{is}=1\} = \frac{Pr\{s-1 \text{ failed inquiries} \cap \chi_{is}=1\}}{\sum_{1\le s'\le \tilde{S}} Pr\{s'-1 \text{ failed inquiries} \cap \chi_{is'}=1\}} \tag{12}$$

Using the linearity of the integral operator, the denominator may be written as

$$\begin{aligned} &\lambda \int H(r^*) \sum_{1\le s'\le \tilde{S}} \left(1-p_h H(r^*)\right)^{s'-1} dF_h(r^*) \\ +\ &(1-\lambda)\int H(r^*) \sum_{1\le s'\le \tilde{S}} \left(1-p_l H(r^*)\right)^{s'-1} dF_l(r^*) \end{aligned}$$

Letting $\tilde{S}$ go to infinity and substituting back into 12 yields the likelihood contribution of an application with s

inquiries:

$$l(S_i = s|\chi_{is} = 1) = \frac{Pr\{s-1 \text{ failed inquiries} \cap \chi_{is} = 1\}}{\lambda/p_h + (1-\lambda)/p_l} \tag{13}$$

where the numerator is defined as in equation 11. Combining this with the likelihood of each realized loan from equation 10 yields the likelihood for our full data.[43]

Although well-defined, maximizing the likelihood defined by the above equations remains difficult. Given two joint distributions, we must estimate five parameters associated with the type distribution, and default and acceptance probabilities, as well as three distributions: the offered rate distribution $H(o)$, and the reservation rate distributions for high and low types, $F_h(r^*)$ and $F_l(r^*)$, respectively. To ease the estimation burden, we make two simplifying assumptions. First, we assume that high and low type borrowers draw their search costs from the same distribution $G(c)$. This assumption guarantees that the reservation rate distribution for each type is entirely determined by the distribution of search costs and offered rates. To see this, recall that a type z borrower has the following relationship between their search cost c and reservation rate r^*

$$c \quad = p_z \int_{-\infty}^{r^*} (r^* - r)\, dH(r) \quad \equiv \psi_z(r^*)$$

That is, we may express a borrower's of type z's search costs as a monotone function of their reservation rate $\psi_z(r^*)$. Since $\psi_z(r^*)$ is strictly increasing over its domain, its inverse $\psi_z^{-1}(c)$ exists and is strictly increasing. This implies that the distribution of reservation rates for type z individuals may be expressed as

$$F_z(r^*) = G\left(\psi_z(r^*)\right)$$

In addition, letting $g(c)$ be the pdf of the search cost distribution, and $f_z(r^*)$ the pdf of the reservation rate distribution for type z individuals, we may write

$$f_z(r^*) = g\left(\psi_z(r^*)\right) \frac{d\psi(r^*)}{dr^*}$$

If $\psi_z(r^*)$ is easily calculable, then estimating the distribution of borrower search costs and offered rates is sufficient to estimate the distribution of reservation rates for each type of worker. This greatly simplifies the estimation problem: rather than estimate three distributions, we now only require two. To feasibly calculate the mapping between search costs and reservation interest rates $\psi_z(r^*)$, we impose our second assumption: that the offered rate distribution is well-approximated by a mixture of N normally distributed random variables parameterized by $\beta_H \equiv \left\{\mu_H^{(n)}, \sigma_H^{(n)}, \pi_H^{(n)}\right\}_{n=1}^{N}$, while the search cost distribution is well-approximated by a mixture of N log-normally distributed random variables parameterized by $\beta_G \equiv \left\{\mu_G^{(n)}, \sigma_G^{(n)}, \pi_G^{(n)}\right\}_{n=1}^{N}$. That is, we assume that we may write

[43] We do not observe the universe of realized loans. We therefore assume that the probability of observing any given loan is independent of all other events, and thus is additively separable in the log-likelihood function.

$$h(r) \approx \sum_n \pi_H^{(n)} \frac{1}{\sigma_H^{(n)}\sqrt{2\pi}} \exp\left[-\frac{\left(r-\mu_H^{(n)}\right)^2}{2\left(\sigma_H^{(n)}\right)^2}\right] \qquad g(c) \approx \sum_n \pi_G^{(n)} \frac{1}{c\sigma_G^{(n)}\sqrt{2\pi}} \exp\left[-\frac{\left(\log c-\mu_G^{(n)}\right)^2}{2\left(\sigma_G^{(n)}\right)^2}\right]$$

for $\pi^{(n)}$ the mixing weight on the n^{th} normal distribution, $\mu^{(n)}, \sigma^{(n)}$ the mean and standard deviation parameters of the n^{th} underlying normal distribution. This assumption permits the analytical construction of the reservation rate distribution for high and low type individuals, and is motivated by the roughly normal distribution of residualized realized rates observed in Figure 1. A detailed description of this construction is provided in Appendix E.1.

To estimate our parameters, we maximize the log likelihood for our sample of loans and applications. We assume that an approved loan application is reported in our loan-level dataset with i.i.d. probability q. We consider q to be a nuisance parameter whose estimation is not of interest. Let the set of observations in the realized loan dataset be given by $\mathscr{L}$, while the set of observations in the application dataset be given by $\mathscr{A}$. We therefore maximize the following log-likelihood with respect to a choice of $\theta \equiv \{p_h, p_l, x_h, x_l, \lambda, \beta_H, \beta_C\}$

$$L(\theta; q) = \sum_{i\in\mathscr{L}} [\log q + \log l(R_i, D_i, S_i | A_{is}=1, \chi_{is}=1, \theta, t, T)] + \sum_{i\in\mathscr{A}} [\log(1-q) + \log l(S_i|\chi_{is}=1;\theta)]$$

where $l(R_i, D_i, S_i|A_{is}=1, \chi_{is}=1, \theta, t, T)$ is given by equation 10, and $l(S_i = s|\chi_{is}=1, \theta)$ is given by equation 13. Since q is additively separable from θ, its value will not affect our optimal choice of $\hat{\theta}$. To uniquely identify the parameters, we impose that $p_h \geq p_l$, but impose nothing about the relationship between x_h and x_l.

To prepare the data for estimation, we residualize all observed interest rates to reflect information that the lender can observe about the borrower without an in-depth screening. Following equation 2, we regress origination interest rates on the borrower's sex, race, age group, education, income group, and debt-to-income group, as well as origination year and property state fixed effects. As a result, our estimates should be interpreted as allowing lenders to discriminate along easily observable characteristics based on price. Second, we winsorize all applications with more than 11 inquiries, in order to match the maximum number of inquiries observed in the realized loan dataset.

D.2 Calculating Market Shares

To construct the market share of type z individuals as a function of a bank'soffered interest rate $q_z(r)$, consider the probability that a type z borrower with reservation rate r^* borrows at rate $R \leq r$. If $r^* \leq r$, this probability will be 1, as the borrower will never apply for a mortgage at a rate above r. Suppose now that $r < r^*$. Since search is undirected and the application approval process is independent of the search process conditional on a borrower's type, this probability is equal to the probability that the borrower is offered a rate less than or equal to r, given that she was offered a rate less than r^*. Thus,

$$Pr\{R \leq r | r < r^*\} = \frac{H(r)}{H(r^*)}.$$

Let $F_z(r^*)$ and $f_z(r^*)$ be the distribution and density, respectively, of type z reservation rates. Integrating out the condition on the borrower's reservation rate yields the share of the type z market accounted for by lenders charging a rate less than r

$$Pr\{R \leq r | Z = z\} = \int_r^\infty \frac{H(r)}{H(r^*)} f_z(r^*) dr^* + F_z(r).$$

Taking the derivative of the above equation with respect to r yields the market share of lenders charging a rate r:

$$\frac{dPr\{R \leq r | Z = z\}}{dr} = \int_r^\infty \frac{h(r)}{H(r^*)} f_z(r^*) dr^*$$

Finally, since a mass $h(r)$ of lenders charge interest rate r, and the borrower samples each of these lenders with equal probability, the residual demand curve for a lender charging rate r is the above quantity divided by $h(r)$:

$$q_z(r) = \int_r^\infty \frac{f_z(r^*)}{H(r^*)} dr^*$$

as in equation 5. Taking the derivative of the above expression yields the downward slope of the residual demand curve from type z individuals, reflecting the market power that the search process gives banks:

$$\frac{dq_z(r)}{dr} = -\frac{f_z(r)}{H(r)} < 0. \tag{14}$$

D.3 Estimating The Cost of Making a Loan

In order to construct robust counterfactual analyses, one must impose structure on the determination of equilibrium offered rates in the market. We thus estimate the cost of making loans in the market. Recall that, as in section 5.3, lenders choose offered rates r in order to maximize expected profits. All lenders share a common cost of making a loan m. Let borrower creditworthiness x_z reflect the probability that the borrower never defaults on her loan. We assume that a borrower defaults at a constant hazard, so that the probability that a type z borrower with loan of term T survives through t periods is $x_z^{t/T}$. This implies that a bank will expect to reclaim a fraction $\tilde{x}_z = (x_z - 1)/\log(x_z)$ of every dollar loaned to a type z borrower.[44] As a result, letting S be the size of the market, the expected profits from making a loan at rate r are

[44]To see this, suppose a borrower originates a mortgage whose term is T, requiring N discrete payments of equal size. Letting $\Omega(t)$ be the survival function after a fraction t of the loan's life, we have that the expected repayment is $\sum_{1 \leq n \leq N} \Omega(nT/N)/N$. Substituting in for $\Omega(t)$ using the proportional hazard assumption implies that the expected repayment can be expressed as

$$\frac{1}{N} \frac{x_z^{\frac{1}{N}}(1 - x_z)}{1 - x_z^{\frac{1}{N}}}.$$

Taking the limit as N tends to infinity yields the result.

$$\mathbb{E}[\Pi(r|m)] = S\left[\lambda q_h(r)\left(r\cdot\left(\frac{x_h-1}{\log(x_h)}\right)-m\right)+(1-\lambda)q_l(r)\left(r\cdot\left(\frac{x_l-1}{\log(x_l)}\right)-m\right)\right]$$

where $q_z(r)$ is given by equation 5. The adverse selection problem presents a challenge for standard first order approaches to maximization and implies that certain observed rates are difficult to rationalize. To match the data, we thus exploit the fact that most mortgage rates are offered according to increments of 1/8 of a percent. Following the logic of section 5.3, we transform the interest rate setting problem into a discrete choice problem, in which lenders choose from a menu of K discrete potential rates to offer. This approach leads to the offered rate choice probabilities expressed in equation 6:

$$Pr\{j \text{ choose } r_k|m,\sigma_\xi\} = \frac{\exp\left(\mathbb{E}[\Pi(r_k|m)]/\sigma_\xi\right)}{\sum_{\tilde{k}=1}^{K}\exp\left(\mathbb{E}[\Pi(r_{\tilde{k}}|m)]/\sigma_\xi\right)}$$

In equilibrium, this offered rate distribution must be consistent with the offered rate distribution $H(o)$ used to calculate the market shares expected from choosing rate r, as determined by 5. Furthermore, the maximum likelihood estimates of $H(o)$ must align with these choice probabilities. This suggests a robust approach to estimating the supply side parameters by minimizing the distance between our maximum likelihood estimates of $H(o)$ and the choice probabilities as given by equation 6. Specifically, we choose the cost of making a loan m in order to minimize the distance between the mean and variance of the maximum-likelihood implied offered rate distribution, and the logit-choice probability distribution.

E Computational Details

E.1 Constructing Reservation Rate Distributions from Search Cost Distributions

Since high and low type borrowers draw their search costs from the same distribution $G(c)$, recall that one may express a borrower of type z's search costs as a monotone function of their reservation rate:

$$c \quad = p_z \int_{-\infty}^{r^*} (r^*-r)\, dH(r) \quad \equiv \psi_z(r^*)$$

In this section, we derive analytical expressions for $\psi_z(r^*)$ under the assumption that the distribution of offered rates and search costs are well approximated by a mixture of normal and log-normal distributions, respectively. That is, we assume that we may write

$$h(r) \approx \sum_n \pi_n^H \frac{1}{\sigma_n^H\sqrt{2\pi}} \exp\left[-\frac{\left(r-\mu_n^H\right)^2}{2\left(\sigma_n^H\right)^2}\right]$$

for π_n^H the mixing weight on the n^{th} normal distribution, μ_n^H, σ_n^H the mean and standard deviation parameters of the n^{th} underlying normal distribution. Similarly, we assume that the search cost distribution is well approximated by a mixture of log-normal distributions parameterized by $\beta_G \equiv \{\mu_n^G, \sigma_n^G, \pi_n^G\}_{n=1}^N$.

Suppressing the superscript H on the parameters of the normal mixture for presentation, and letting $pdf_{\mathcal{N}(\mu,\sigma)}(x)$ and $cdf_{\mathcal{N}(\mu,\sigma)}(x)$ be the pdf and cdf of a normal distribution with mean μ and standard deviation σ evaluated at x, we have:

$$\begin{aligned}
\psi(r^*) &= p_z \int_{-\infty}^{r^*} (r^* - r)\, dH(r) \\
&= p_z r^* H(r^*) - p_z \sum_n \pi_n \int_{-\infty}^{r^*} \frac{r}{\sigma_n \sqrt{2\pi}} \exp\left[-\frac{(r-\mu_n)^2}{2\sigma_n^2}\right] dr \\
&= p_z r^* H(r^*) - p_z \sum_n \pi_n \left[\mu_n cdf_{\mathcal{N}(\mu_n,\sigma_n)}(r^*) - \sigma_n^2 pdf_{\mathcal{N}(\mu_n,\sigma_n)}(r^*)\right]
\end{aligned}$$

where the third equality follows by integration by parts. The above expression may be numerically inverted in a computationally-efficient way. Also observe that we may calculate the derivative of $\psi_z(r^*)$ to be

$$\frac{d\psi(r^*)}{dr^*} = \frac{d}{dr^*}\left[p_z \int_{-\infty}^{r^*} (r^* - r) dH(r)\right] = p_z H(r^*)$$

which is easy to compute given our approximation to $H(o)$. Thus we may construct the distribution of reservation rates for a type z individual given our approximation of $G(c)$ and $H(o)$.

E.2 Computing Counterfactual Offered Rate Distributions using Lenders' Profit Maximization

Changing any of our parameters will change the equilibrium distribution of rates offered in the market. Adjusting the search cost distribution or probability that an application is accepted changes the reservation rate distributions which enter into the market share equations (5) and (14). Meanwhile, changes to λ, x_h, m, or x_l directly impact the relationship between lender loan costs and their optimally-offered rate. Counterfactual analysis therefore necessitates a method of computing counterfactual offered rate distributions that constitute Nash equilibria.

Since both the market share equations (5) and (14) and reservation rate distributions depend on the distribution of offered rates in the market, a lender's optimal offered rate choice $\hat{r}$ will depend on the choices of all other firms in the market $H(r)$. In equilibrium, the distribution of offered rates implied by the lenders' profit maximization problem $\hat{H}(\hat{r})$ must be the same as the distribution of rates $H(r)$ used to calculate a lender's market share functions. Thus we need to solve a functional fixed point problem for $H(r)$.

Our approach proceeds in three steps. First, we guess a normally-distributed equilibrium offered rate distribution $H(r; \beta_H)$. Next, we use equation 6 to calculate an implied distribution of optimally-offered rates $\hat{H}(r; \beta_H)$. Finally, we minimize the distance between $H(r; \beta_H)$ and $\hat{H}(r; \beta_H)$ with respect to β_H. The problem may then be written as

$$\min_{\beta_H} \quad ||H(r;\beta_H) - \hat{H}(r;\beta_H)|| \tag{15}$$

for some appropriately chosen norm $||$. We solve this problem using numerical gradient-descent optimization algorithms implemented with KNITRO, and match the mean and variance of the implied distributions to that of the guessed distribution.[45] Once the equilibrium distribution of offered rates is calculated, it is straightforward to produce counterfactual simulations of the demand side of the model.

This approach faces two potential problems. First, multiple equilibria may arise, as changes in the offered rate distribution endogenously determine borrowers' reservation rate strategies, which in turn affect the optimal offered rate distribution. To address this issue, we experiment with multiple starting values when searching for equilibria with the approach laid out above. Across all of our starting values, we find the same equilibrium offered rate distributions.

A second concern arises from numerical approximations. We approximate the equilibrium offered rate distributions with normal distributions, which are then fed into the market share equations in order to calculate logit choice probabilities for every feasible rate. The objective function in the minimization problem 15 therefore compares a normal distribution with logit-implied choice probabilities, which will naturally involve some error. To evaluate the severity of this concern, we search for an equilibrium using the set of parameters estimated using our maximum likelihood routine. The mean and standard deviation of the MLE offered rate distribution are 0.142 and 0.547, respectively. By comparison, the "equilibrium distribution," obtained by running these parameters through the equilibrium search routine described above has a mean and standard deviation of 0.206 and 0.723, respectively. Although imperfect, we consider this error to be relatively small. After simulating the demand side of the model, this leads to a gap in average rates paid of 2.9bp, and an increase in search of 0.13 inquiries per borrower. For all counterfactuals in which we allow the offered rate distribution to adjust, we compare the counterfactual output against "equilibrium" simulations, which are based on a normally-distributed offered rate distribution with mean and standard deviation of 0.206 and 0.723, respectively.

[45] It is unnatural to assume that offered rates will be well-approximated by a single normal distribution under the redlining counterfactual. In this counterfactual, we therefore approximate the offered rate distribution with a mixture of two normal distributions - one for redlining lenders and another for non-redlining lenders - and find an associated logit-implied distribution for each. Our objective function then minimizes the weighted sum of the distance between each normal and logit-implied distributions.