Searching for Approval

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

Preliminary: Please do not cite without authors' permission

November, 2018

Abstract

We study the interaction of search and approval process in credit markets. We use a unique dataset that details search behavior for a large sample of mortgage borrowers, and loan application and rejection decisions. Our data reveal substantial dispersion in mortgage rates and search intensity. 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. We provide evidence that this occurs because lenders screen borrowers' creditworthiness, rejecting unworthy borrowers. This approval process 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. The model rationalizes the facts in the data and reveals that search behavior is determined not only by consumer sophistication but also by the approval process that relies on the underlying distribution of borrower quality. We estimate the parameters of the model and study several counterfactuals. We find that using "overpayment" by consumers as a proxy for consumer unsophistication may be misplaced since it partly represents rational search in presence of rejections. Moreover, presence of better models with big data, could endogenously lead to more severe adverse selection in credit markets. Finally, place based policies, such as the Community Reinvestment Act, may lead to higher prices in equilibrium that reflect the endogenous search response rather than increased credit risk.

*We thank seminar participants at the University of Chicago, UT Austin and Stanford GSB for useful comments.

1 Introduction

Consumer credit markets exhibit substantial price dispersion. In mortgage markets, for example, borrowers with similar characteristics obtain mortgages with substantially different interest rates or fees (Gurun et al 2016, Allen et al. 2014, Hall and Woodward 2012). Moreover, the high cost of credit for less educated, poor, low credit score (subprime), or minority borrowers are ascribed to the lack of their financial sophistication. A leading explanation of these facts is consumer search. Less sophisticated borrowers search less, and consequently find worse, more expensive financial products. 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 in financial markets and outside mainly infers search behavior from the price distribution, or in rare cases measures search behavior in 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 dataset of conforming mortgages from a large government sponsored entity (GSE) in the United States. Matching the data with consumer credit reports from a large national credit bureau permits a unique look at search behavior of borrowers, combined with detailed information on borrowers characteristics, and application acceptance decisions. These data allow us to explore search in substantially more detail than the literature has done heretofore.

In order to understand search behavior in the mortgage market, one has to acknowledge a central distinction between search in credit markets and markets for non-financial consumer goods. In the mortgage market lenders evaluate borrowers' creditworthiness, and approve or reject customers based on this evaluation. If an application is rejected, customers have to apply for a mortgage with another lender. This approvals process differentiates search in the mortgage market from search in the goods market, such as books or auto market. In fact, most empirical search models, which are applied to financial markets do not draw a distinction between searching for financial services and searching for consumer goods.[1] This type credit 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. A similar approval process takes place in the insurance industry, where applicants are screened for underlying risks.

Since few studies can measure search directly, we first document several facts related to mortgage search, pricing, and mortgage approvals. Consistent with previous research, we find substantial dispersion in mortgage rates paid by borrowers, even after accountting for detailed borrower, loan, time, lender, and location characteristics. These differences in rates result in some borrowers paying thousands of dollars more per year than similar borrowers at the same location, at the same point in time. The median borrower who obtains a mortgage does not search much, having only 2 formal credit inquiries around the mortgage approval on her record.

We then turn to the central fact in the paper. In Figure 1 we plot average origination rates on mortgages for borrowers with different amount of search. The average realized interest rate does not monotonically decline with search. Borrowers, who search a lot, obtain worse mortgages than borrowers, who search little. The fact that mortgage rates do not decline monotonically with search is very robust, and survives across different subsamples of borrowers, after extensive controls for borrowers' characteristics, and after conditioning both on location as well as

[1] See, for example, Gurun et al 2016, Allen et al. 2014, Hall and Woodward 2012.

time of borrowing. As we discuss later, this result cannot be generated by canonical search models, which have been applied to the market for books, mutual funds, and to the mortgage market as well .

Second, using loan performance data, we show that loans from borrowers who search are more likely to be delinquent or default ex post, even conditional on very detailed ex ante measures of their creditworthiness, such as FICO, LTV, DTI, which we obtain from the lender's. In other words, borrowers who search more, all else equal, are more likely to become delinquent or default on their loans. Third, we show a robust negative relationship between the probability of mortgage approval and the number of searches. In other words, borrowers, who search more are more likely to have their mortgage application rejected. This is the case even in a narrow time window during which prior borrower searches are not observed by the lender.

To understand these facts, and to better understand the interaction between search and approvals, we develop a search model of 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 their creditworthiness is private information. 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 can choose to search for another lender, incurring her search cost once more.

The approval process affects borrowers' search, because they have to 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 mortgage to avoid future search. In other words, because of the possibility of rejection, low creditworthiness borrowers' search behavior will mimic that of financially unsophisticated, high search cost borrowers. This simple intuition has several implications.

First, one cannot infer consumers' search costs from the prices they pay. 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 argue that "overpayment" by certain types of borrowers can be interpreted as a sign of their low financial sophistication. 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. So, for example, if minority borrowers pay higher rates than non minority borrowers, all else equal, this could be a consequence of low sophistication of minority borrowers, or, as our model suggests, a rational response of minority borrowers to higher rejection rates.

Second, the approval process endogenous adverse selection. Moreover, adverse selection arises even when high creditworthiness borrowers have a higher preferences for loans, which generally results in advantageous selection. Since low creditworthiness borrowers behave as if they have higher search costs, such borrowers are endogenously more likely to take up expensive mortgages. Thus adverse selection arises endogenously, even if underlying preferences

suggest advantageous selection.

We estimate the model and show that it successfully replicates the qualitative and quantitative relationship between mortgage pricing, approval, and search that we find in the data. The model is consistent with three sets of facts we observe in the data, which the canonical search model has a difficult time explaining. First, it explains why borrowers who search a lot pay higher interest rates on average. These are borrowers with low creditworthiness, whose 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 if they are approved. In other words, they pay high rates because of their search behavior, not their low creditworthiness.

Second, we our model can explain the relationship between search, default, and loan approvals. Standard search models suggest that the relationship between interest rates and search is solely driven by search cost. They have no room for for rejecting a mortgage application, and suggest search and default are independent. Our model suggests that borrowers who search a lot are of low creditworthiness. If that is indeed the case, when their quality is revealed ex post in repayment behavior, frequent-searchers should be more likely to default. Consistent with the prediction, borrowers, who search a lot, are more likely to be delinquent and default on their loans ex post. Second, the model predicts that less creditworthy borrowers are more likely to be rejected because the information is partially revealed after they undergo screening by the lender. Using novel data on mortgage approval, we show a robust negative relationship between the probability of mortgage approval and the number of searches. 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. These borrowers, with approval rate of almost 98.75%, differ substantially from the overall population, whose rejection probability is approximately 18%. The subsample of rarely-rejected borrowers is interesting, because our model predicts that the correlation between search and mortgage rates should be negative for this specific subpopulation. Rarely-rejected borrowers should sort only on search costs, so borrowers who search more obtain cheaper mortgages. Note that this prediction is in stark contrast to our estimates for the overall population of borrowers. Strikingly, we do find that, in our population of rarely-rejected borrowers, mortgage origination rates are monotonically decreasing in the frequency of search. These results provide additional support for our model, and suggest that the non-negative relationship between search and mortgage rates for the overall sample is indeed driven by the approval process rather than some other unobservable borrower characteristic.

In order to pursue interesting counterfactual analyses, we next estimate the model. We employ a maximum likelihood approach using data on the joint distribution of search, origination rates, application approvals, and default. Consistent with intuition, we find that riskier populations, as measured by low FICO scores and high loan-to-value (LTV) ratios, are more likely to have their application rejected, inducing higher prices among these groups.

Using the estimated model, we first consider the impact of tightened lending standards of the sort seen during the

financial crisis. Our model shows that lenders' reduced willingness to lend to borrowers not only reduces borrower access to credit, but increases both search and the prices paid on loans. Because borrowers internalize the tighter lending standards into their reservation price, they are willing to accept more expensive loans. A decline in application acceptance probability of a magnitude similar to that in the crisis raises the average rates paid by borrowers by 0.8 basis points (bp), absent any change in the distribution of rates posted by lenders. Furthermore, this increase in reservation rates induces lenders to increase their offered rates, pushing rates yet higher. With this supply side response, we estimate that tighter lending standards during the crisis increased average mortgage rates by 25.4bp.

Next, we pursue two counterfactual exercises to address the question of discrimination. First, we show that the practice of redlining - in which a subset of lenders selectively reject a large portion of some discriminated population - is 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. This effect arises because such discriminated groups internalize the increased rejection probability into their reservation rates. Our estimates imply that if half of the lenders in a region rejected borrowers at twice the rate of non-redlining lenders, average realized mortgage rates increase by 28.7bp. Although this rise is concentrated amongst the discriminated group, those not belonging to the discriminated group also suffer higher rates, due to the strategic complementarity in bank rate setting.

Second, we study the impact of policies such as the Community Reinvestment Act (CRA), which impelled lenders in particular locations to increase their application acceptance probabilities for all borrowers. Specifically, we consider a counterfactual exercise in which the CRA renders screening uninformative, so that borrowers of both high and low creditworthiness are rejected at the same rate. Absent any supply side response, we see that average rates in the market drop by 2.5bp for low creditworthiness borrowers in accordance with their reduced reservation rate. However, when we allow lenders to adjust the rates they offer to the market, the mean rate falls by a further 26.4bp.

Overall, our results suggest that search in credit markets differs substantially from search in other product markets. When selling a car, book, or toothpaste, the seller's payoff does not depend on the identity of the consumer beyond the price she pays for the product. With credit (and insurance) products, the seller's payoff critically depends on the characteristics of the borrower. The standard (informative) credit approval process substantially alters the search incentives of borrowers, and changes which types of borrowers sort to which types of mortgages. This sorting is inconsistent with standard search models, and prevents identification of the search cost distribution from price data alone. Moreover, the approval process leads to endogenous adverse selection, which affects both the search incentives of borrowers, as well as the pricing incentives of the sellers. Accounting for the credit approval process is therefore critical in 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.

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, Hall and Woodward 2012, Alexandrov and Koulayev 2017). The role played by switching costs/consumer inertia in the context of health insurance choices was studied by Handel et al. (2013). In their setting, consumers self-select into a contract from a menu of contracts, as in a number 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 approval process. 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.

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 in detail. Section 3 describes the data used in our empirical analysis in detail. In section 4, we present the basic facts of search in mortgage markets, as well as the relationship between search and prices, delinquency, and application approval rates. 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. We describe and report the estimation of our model in section 7. Finally, section 8 describes and reports the results of our counterfactual analyses. Section 9 concludes.

2 Credit Application Process and Inquiries

The formal process of getting 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. 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.[2] 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 within 45 days of the "final" mortgage application (and approval) to capture the intensity of borrower search. Therefore it is useful to discuss several details related to inquiries and search in the mortgage market. First, it is possible that borrowers search for mortgages informally without a credit pull, for example, by searching for 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

[2]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).

the house characteristics. Thus, not being able to measure such informal searches should not impact the manner in which one should think about borrower search.

Second, similar formal inquiries might 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 inquiries in the credit bureau data. Would these non-mortgage inquiries then conflate the "total inquiries" that we treat as mortgage search? Several observations suggest the answer is no. To start with, 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 lowered when borrowers take up credit products, borrowers have strong incentives not to formally search for other credit products such as credit cards before applying for a mortgage.

We also formally check whether non-mortgage inquiries pollute total inquiries in two ways. One, using credit bureau data merged with approved loan information, we measure the share of mortgage-related inquiries[3] as a proportion of total inquiries for a given borrower in the one month prior to her mortgage origination. The one month window reflects 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 mortgage shopping period over which we examine inquiries.

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 mortgage applications from 2001 to 2013 that are used to purchase or refinance a single family property. 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. 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

[3] As determined by the credit bureau.

property location, for both approved and rejected loan applications.

Our second dataset contains approximately 1.3 million mortgages that are approved and originated between 2001 and 2011. At origination, we observe the borrower's credit score, the loan-to-value (LTV) ratio, the loan characteristics (origination balance, note rate, and term), the backend ratio, whether the loan was originated through a broker, loan purpose, occupancy, and the location of the mortgaged property (zip code, city (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. 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.[4]

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 FICO score of 725.8 substantially exceeds that of the US population, which was 688 in April 2011,[5] 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, borrowers were slightly less creditworthy in the applications sample, with average FICO of 707.4, and average CLTV of 75.3%. This difference suggests that less creditworthy borrowers face a lower probability of their mortgage applications being accepted. There is substantial creditworthiness heterogeneity 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: 15.95% had entered default by January, 2015.

Our dataset includes loans originated throughout the crisis period. 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] To limit the influence of outliers, we winsorize applications and loans lying above the 99th percentile of inquiries, interest rates, DTI, or LTV ratios.

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

4 Price Dispersion and Differences in Search

Differences in mortgage rates across borrowers have frequently been attributed to costly search. 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 (Gurun et al 2016; Allen et al 2014). Borrowers pay substantially different mortgage rates in our sample as well, even after adjusting for points and fees. We present the full distribution of rates across three origination time periods in Figure 2A, showing substantial rate dispersion. Figure 2B 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 might arise because of borrower differences. 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 estimate the following specification:

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

in which r_{itm} represents the origination rate of borrower i at time t in market m. X_i are the borrower's 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, 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 market fixed effects,[6] μ_m, and on time fixed effects μ_t. Our data set was expressly collected by the lender for the purposes of making the loan, so these controls 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. Mortgages with negative (positive) residuals are cheaper (more expensive) than the mean mortgage with the same characteristics. The distribution of these residuals (Figure 2C) 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

[6] We define a market to be a state.

$1,140 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.[7] 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 two basic facts related to borrower search. First, there are differences in search amounts across borrowers. The second fact we document is that borrower characteristics, which are generally associated with consumer sophistication, do 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 eventually obtain a mortgage. The median borrower who obtains a mortgage does not search much, having only 2 inquiries on her record (Figure 3). In fact, a borrower in the 75^{th} percentile searches 3 times. Mortgage applicants search substantially more, with a median of 9. 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.

Borrower characteristics such as education, income, age, and race have been used as proxies for consumer sophistication in the literature (Hall and Woodward 2012, Gurun et al 2016). Sophisticated consumers should have lower search costs, and therefore search more. Consider differences in search across FICO levels in Figure 3C, and across educational attainment in Figure 3D. 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).[8] 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:

$$s_{itm} = \alpha + \beta X_i + \mu_m + \mu_t + \varepsilon_{itm} \tag{1}$$

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. We examine the conditional correlation between search and borrower characteristics, such as

[7] 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 R^2 from 0.80 to 0.81 and reduces the standard deviation from $41bp$ to $40bp$.

[8] 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.

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 and time fixed effect μ_m and μ_t. Any differences in the regulatory environment are also absorbed by the location fixed effect. We present the results in Tables 3 and 4. 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 the argument, more creditworthy borrowers search less, even conditional on other characteristics, suggesting an important role for creditworthiness in understanding consumer search behavior.

4.3 Do Borrowers who search more obtain cheaper mortgages?

The benchmark search model suggests that search and transacted prices are negatively correlated, as we more formally illustrate in Section 5.5.1. Intuitively, low search cost (financially savvy) consumers find searching cheap. This low search cost allows them to search more, and find better, cheaper products. 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 worse, more expensive products on average.

We first present a simple cut of the data by plotting the average mortgage rate as a function of search in Figure 1. Under the benchmark model, the average price (origination rate) should monotonically decline with search. Figure 1, 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 searching is correlated with increased mortgage rates. High-inquiry borrowers, who search a lot, obtain worse 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.

We cut 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 in Figure 4 and Appendix Figure 18. We find the same pattern for low, middle and high FICO scores, low, middle and high educated populations, for black, white, and Hispanic borrowers, as well as for low, middle, and high income borrowers. 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.

To show that our results are indeed robust, 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

$$r_{itm} = \alpha + \sum_{s=2} \beta_{s'} \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + \beta X_i + \varepsilon_{itm} \tag{2}$$

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, such as the borrowers FICO score, their loan to value ratio (LTV), race, income, and others. To ensure that our results are not driven by local supply or demand conditions, we include the time fixed effect μ_t and location fixed effect μ_m. These fixed effects will also absorb any aggregate fluctuations, such as changes in the risk premia, or persistent differences across markets.

In effect, we consider two borrowers in the same location, at the same point in time, with the same FICO score, income, race, and other characteristics observed by the lender, and compare how the interest rate charged on their mortgage differs with the amount of search. We plot the coefficients β_s in Figure 5. As the figure suggests, borrower, location, or time differences do not drive our result. Increased search has a U-shaped, or even monotically increasing relationship with interest rates. We next show that the results persist across different sub-populations. First, we cut the data by borrower creditworthiness (FICO), which is strongly correlated with both mortgage rates and search. We split the sample into three different FICO populations, and estimate specification 2 for each of them. Figure 5 plots the estimates. If anything, the results are even more striking than the baseline. As in Figure 4, 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. We present the results in Table 7. 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. 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

In this section, we present a model which can rationalize the observed U-shaped or positive relationship between search and realized prices in the mortgage market. We 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, it permits a deeper understanding of search in markets of asymmetric information and approvals. Second, the model yields testable predictions that distinguish it from standard search models, which we test in section 6. Third, the model is both tractable and realistic enough to be estimated, and used to conduct policy-relevant counterfactual analyses in Section 8.

Our model is an extension of the standard sequential search model first proposed by Carlson and McAfee (1983); indeed, given a set of parameters which trivialize the application approval process, 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. Unlike in standard search models, mortgages 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.

5.1 Setting

5.1.1 Borrowers

Consumers are indexed by iz and have two characteristics, search cost $c_i \sim G(c)$, and repayment ability $x_z \in (x_h, x_l)$, with $Pr(x_z = x_h) = \lambda$. Borrowers with high repayment ability (creditworthiness), x_h are more likely to repay a loan than borrowers with low repayment ability: $x_h > x_l$.[9] Creditworthiness and search costs are $i.i.d$ across consumers and types.[10] A consumer iz's utility from obtaining a mortgage from lender j at rate $r_j > 0$ 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.

5.1.2 Lenders and Mortgage Approval

Lenders post mortgage interest rates. Lenders choose from a menu of K discrete potential rates to offer, $r_k \in \{r_1, \ldots, r_K\}$.[11] Lender j's expected profit on a loan to type z at rate k is:

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

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 for the lender, 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 approval

[9] We provide some empirical evidence that two types are sufficient in capturing most richness in the data in Section 12

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

[11] 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.

process. The lender carries out an in-depth check of applicants' creditworthiness, which generates an informative, but imperfect, signal $s_i \in (s_h, s_l)$.[12] If the borrower is of repayment ability x_z, the probability that she is revealed as such is $p_z = Pr(s_h|x_z)$. The in-depth review is informative $p_h \geq p_l$, so high repayment ability borrowers are more likely to be revealed as good.[13] 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 can search again.[14]

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:

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

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

[12] In an alternative specification, we allow lenders to choose whether or not to screen applicants at a constant cost $\gamma \geq 0$. Such a model generates identical predictions. Furthermore, estimating this version of the model reveals that γ is very close to 0 so that all lenders would choose to screen at no cost. In what follows, we therefore present the model without screening costs in order to focus on the key mechanisms at work.

[13] We assume that applicants receiving the high signal s_h are approved for a mortgage, while those who received the low signal s_l are not.

[14] 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.

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

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.

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. Because low-quality borrowers are less likely to be approved for a mortgage, $p_l < p_h$, they behave as though they have higher search costs, and are willing to accept worse mortgages. Formally, consider two borrowers with the same search costs, but different creditworthiness. Then:

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

$p_h > p_l$ implies that $r^*_{ih} < r^*_{il}$. That is, less creditworthy borrowers are willing to accept higher rate mortgages 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.[15]

To better illustrate the adverse selection problem, we present a numerical example. Figure 7A shows the differences in reservation interest rates for high and low creditworthy types with the same search 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, also apply for expensive mortgages, because they understand that the chances of mortgage approval are low in the future. Figure 7B 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

[15] Conversely, adverse selection arises even if high quality borrowers value mortgages more, i.e. if $\sigma > 0$. Intuitively, adverse selection in this model occurs on the intensive margin: all borrowers will find a mortgage in the limit. The overall preference for mortgages captured in σ operates on the extensive margin of obtaining a mortgage in the first place, and therefore drops out of the search problem.

high reservation rates. Differences in approval rates across types therefore lead to adverse selection in the mortgage market.

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 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, as well as the probability of repayment on the pool of mortgages. For every mortgage, the expected profit depends on the lender's market share of the two types of borrower based on the positive signal and the rate offered, as well as the cost of funding, m. 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:[16]

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

Appendix 13.2 shows that the market share of type z individuals that a bank offering rate r earns may be expressed as

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

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^*. Each lender faces an additional idiosyncratic profit shock to charging specific rates $\xi_{j,k}$, which are i.i.d and distributed Type 1 Extreme Value (T1EV) with scale factor σ_ξ.[17] 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:

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

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. Put another way, any difference in firms' cost base or regulatory

[16] The profit function is specified in terms of percentage points of interest. In our empirical application, we residualize observed interest rates against borrower characteristics in our empirical analysis, 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 equivalent borrower in the market.

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

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:

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

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. 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. Recall that if the approval process is uninformative $p_h = p_l$, the model reduces to the benchmark model without approvals. 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 14.2.

5.5 Model predictions

In this section, we show how the introduction of private information and an approval process into a standard search model yields several predictions, 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.[18] We show that this benchmark model's predictions are inconsistent with the relationship between search and rates we document in Section 4.3. In this benchmark, 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:

$$c_i = \int_{\underline{r}}^{r_{iz}^*} \left(r_{iz}^* - r\right) dH\left(r\right)$$

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, $Pr\left(Search\, again\right) = 1 - H\left(r_{iz}^*\right)$. Then the probability that a borrower with a reservation rate r^* searches more than s times is:

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

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 Figure 6 for a simulated sample of borrowers. Overall, the relationship between average rates and search in the data rejects this prediction, leading us to consider an search model with informative screening.

5.5.2 Introducing informative approvals: 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

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

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, the more creditworthy borrowers disappear from the population of searchers faster than low creditworthy borrowers. To illustrate this, we simulate a search process with highly informative screening. Figure 7C presents the share of high types left in the population at each level of search, for this simulation. 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 appropriate mortgages.

As a result, borrowers' average reservation rate increases with the number of searches. Indeed, Figure 7D shows a positive relationship between search and interest rates for this simulated sample with informative screening. This is in stark contrast to the baseline model of search without approvals. It is however, consistent with the empirical fact documented in detail 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.

5.5.3 Default and approvals

Our model predicts a specific type of equilibrium borrower sorting. As the number of searches increases, the creditworthiness of the borrower pool declines, as shown in Figure 7C. 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 7E 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 exam. As a result, lenders are more likely to reject borrowers who search more, even if they cannot observe the number of searches. Figure 7F 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 decreasing relationship between search and application approval probability
6. 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 informed approvals from a benchmark model without approvals. As we show in Section 4.3, the relationship between search and prices, (prediction 3) is consistent with the approvals model. We now test our model by verifying that predictions 4 through 6 are also observed.

6 Additional Empirical Evidence

6.1 Loan Performance and Search

Our model predicts that borrowers' ex-post search behavior is informative about their underlying creditworthiness. Because less creditworthy borrowers search more in equilibrium, they should be less likely to repay their mortgage. Figure 8 plots the annualized default rate against the number of inquiries on record for all borrowers in our sample.[19] Panel A shows the 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 less creditworthy.

[19] 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 $1 - y = (1 - \tilde{d})^{\tau}$.

High-inquiry borrowers may simply be of lower credit quality on dimensions observable to the lender. Indeed, Figure 3C and Table 3 show that low FICO borrowers do indeed search more. To test whether frequent searchers are more likely to default even conditional on observables, we estimate the following linear regression:

$$d_{itm} = \alpha + \sum_{s'=2} \beta_s \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + \beta X_i + \varepsilon_{itm} \tag{7}$$

in which i indexes the borrower who originates a mortgage in market m at time t. The dependent variable d_{itm} is an indicator for whether the borrower either defaults or is at least 90 days delinquent on their mortgage payments. 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 default probability for borrowers who search s times compared with those who search just once. To ensure that the correlation between search and mortgage rates is not driven by borrower or mortgage characteristics, we extensively control for observable characteristics collected by the lender, such as the borrower's FICO score, LTV ratio (LTV), race, income, and others. Furthermore, to ensure that our results are not driven by local market conditions, we 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 risk premia, or persistent differences in the regulatory environment.

We plot the coefficients of interest, β_s, in Figure 9. Consistent with our predictions, borrowers who search more are more likely to default or become delinquent on their loans, even conditional on observable characteristics. 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 (Figure 9, and Appendix Figures 20, 21, and 22). 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. These borrowers applications are therefore more likely to be rejected following an indepth credit check, even if the past searches are unobserved to lenders. Using our application-level dataset, we are uniquely able to test this implication of our model. 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.

Figure 10A illustrates the strong negative correlation between search and the probability of mortgage approval. This result persists in specific subsamples of our population: Figure 10A is replicated for three groups of borrower FICO score, and across three origination time periods in Figures 10B and 10C, respectively. Borrowers who search more are of lower average quality in two separate datasets and along two dimensions – default and application acceptance probability. To illustrate that the pattern in 10 is robust, we estimate the following linear regression:

$$a_{itm} = \alpha + \sum_{s=2} \beta_s \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + \beta X_i + \varepsilon_{itm} \tag{8}$$

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 of one, 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 of variables observed by the lender, such as the borrowers FICO score, LTV and DTI ratios, among others, and condition on location and time fixed effects to absorb aggregate and persistent differences across time and space. The coefficients of interest are presented in Figure 11. 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 11. The data therefore support the model's prediction that borrowers who search more more are less likely approved for mortgages, conditional on observables.

The benchmark search model in which borrowers differ only in their search cost, would predict no relationship between search and average borrower creditworthiness. It is therefore unable to 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 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.

The non-negative relationship between search and rates in our model is driven by an informative approvals process.[20] Absent the possibility of application rejection our model collapses to the standard sequential search model: the borrowers who search more should, 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,

[20] Recall that random rejections alone do not generate this result.

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 approval rate of this sample is 98.5%, which is substantially higher than the average approval rate of 82.2%, or 89.7% for high (above 720) FICO scores.[21]

Panels A and B of Figure 12 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 for 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 12C. 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 1. To ensure that the negative relation between search and origination rates for rarely rejected borrowers is robust, we next condition on observables. As in Section 4.3, we estimate the following following linear regression:

$$r_{itm} = \alpha + \sum_{s=2} \beta_s \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + \beta X_i + \varepsilon_{itm} \tag{9}$$

in which i indexes the borrower who takes up a mortgage in market m at time t. The dependent variable r_{itm} is the mortgage origination rate. We again include extensive controls, such as the borrowers' FICO score, their loan to value ratio (LTV), race, and as well as a time fixed effect μ_t and location fixed effect μ_m. The coefficients of interest, β_s are presented in Figure 12D. After conditioning on observables, it remains true that rarely rejected borrowers behave as predicted by standard models of search, which our model replicates if $p_h = p_l = 1$. For this group, borrowers who search more borrow more cheaply, ostensibly because their lower search cost translates into lower reservation rates. The results for these borrowers are again in stark contrast to those we document for mortgage borrowers as a whole. These results strongly 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.

7 Model Estimation and Counterfactual Analysis

The model with search and informative approvals we present in Section 5 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

[21] In Appendix Figure 27 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%,

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 in the standard way popularized by Hortaçsu and Syverson (2004). However, all parameters of the model may be identified using of 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 realized loan search distribution is a parallel shift of the application distribution 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 $H(r)$, 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 $H(r)$ 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 two distinct but related 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 comparibility 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, $\{p_h, p_l, \}$, and costs of making loans m, and the variance of the T1EV profit shocks, σ_ξ.[22]

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 $H(o)$ 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

[22]We observe 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.

$H(o)$ and the choice probabilities as given by equation 6. We describe the details of our estimation in Appendix 13 and computational details of the estimation in Appendix 14.1.

7.2 Results

Data Fit: Despite its simplicity, the estimated model matches observed price dispersion and distribution of searches (Figure 13, Panels A and B). The model replicates an increasing relationship between interest rates and search, and interest rates and default documented in sections 4 and 6 (Figure 13, Panels C and D).[23]

Screening Technology and Adverse Selection: Our maximum likelihood estimates are reported in Table 9. Our estimates suggest that most potential borrowers, 73%, are of low type: they default on the full term of the loan 41% of the time and in expectation repay 66 cents of principal on a borrowed 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 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.

First, these estimates suggest that despite the preponderance of bad borrowers in the population of applicants, the rejections of bad borrowers decrease their share in the approved pool substantially: closer to $\frac{1}{3}$ of the population of borrowers at the average originated interest rate are low type. Second, the difference between p_h and p_l of 0.807 suggest that the screening technology is very informative. a simple back of the envelope suggests that the expected loss on a bad borrower applying is lowered by approximately 81% from $23pp$ to $19\% * 23pp = 4.4pp$. 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 8.3, we estimate that the presence of this screening technology increases aggregate lender profits by \$1.2billion per year.

The informative screening technology provides large incentives for adverse selection. Low creditworthiness 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 13E. Adverse selection is most serious for interest rates between the mean and 50bp above the mean. At the mean origination interest rate, the probability of ever defaulting is 0.373, and the derivative of this default rate with respect to the interest rate paid is 0.178. Small increases in the realized interest rate lead to sizable increases in the default probability at the mean realized rate.[24]

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

[24] 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

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

Finally, Figure 14 shows that riskier populations, as measured by low FICO scores and high loan-to-value (LTV) ratios, are less likely to have their application approved, which leads to higher prices among these groups. The share of high types in the population of high FICO borrowers is 8.8 percentage points higher than amongst low FICO borrowers. In addition, low type borrowers with a high FICO score are accepted at with probability 20.8%, compared with just 16.9% for low FICO borrowers.

Search Costs: The mean of the search cost distribution is estimated at 27.2bp.[25] Our estimates of average costs are in line with 27.3bp in Allen et al. (2013), and $29 monthly in Allen et al. (2014) for the Canadian insured mortgage market. The standard deviation of 12.9bp 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). For a 30-year fixed rate mortgage with principal of $170,000 and interest rate of 4% per year this estimate would translate into a monthly payment increase of $27, or an upper bound cost of $9,719 over the term of the loan.[26]

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 be the cost of lending relative to the mean realized interest rate of an average 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. (2013). 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.

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. We then show that the practice of redlining - in which a subset of lenders selectively reject a large portion of some discriminated population - is sustainable in a sequential search equilibrium, and induces borrowers from the discrminated group to pay higher interest rates on average. Finally, we study the impact of policies such as the Community Reinvestment Act (CRA), which impelled 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 10. Note that, 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 14.2.[27]

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

[26] This estimate is an upper bound assuming the mortgage is never refinanced or prepaid.

[27] Throughout our counterfactual analyses, we use our estimated parameters from the full sample, presented in Table 9 as a baseline.

8.1 Tighter Lending Standards

This tightening of lending standards has been at the heart of policy debates for many years. Traditionally, the debate has centered around the trade off between providing consumers access to credit while simultaneously mitigating systematic risks in the banking sector with a focus on the quantity of lending (Dell'Ariccia et al. 2008, Mian and Sufi 2009, Bassett et al. 2014). Famously, Ben Bernanke was declined for a mortgage during his tenure as FED chairman at the peak of the crisis.[28] Our model provides a unique opportunity to understand the implications of this tighter lending standards along a new and crucial dimension: the borrowers' search response to tightening standards, and the implications for mortgage rates. As we show, tightening of lending standards results in higher mortgage rates even if the underlying costs of providing mortgages do not change.

We tighten the lending standards to mimic the changes in mortgage approvals after the crisis, holding all other parameters such as costs of lending fixed, and the industrial organization of the industry fixed. We estimate the change in approval rates during and after the crisis using a logit discrete choice model. We do so to adjust for changes in the pool of prospective prime borrowers (Table 2).[29] 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 reduce the odds ratio of application approval for both high and low types by 21.8%.

Even absent changes in the cost of lending or industrial structure, a tightening of lending standards of the magnitude seen during the crisis has quantitatively important consequences for the rates paid by borrowers. Figure 15A plots the distribution of realized rates given our baseline estimates and in our tighter lending standards counterfactual. The mean rate paid in the market increases substantially by 25.4bp, which is on the order of a discrete increment in the Fed's policy rate, and results in $301 higher payments per year.[30] Tightening of lending standards also exacerbates the distributional consequences of search costs: the standard deviation of realized interest rates increases by 9.0bp. In other words, the additional cost of borrowers who pay rates one standard deviation above the mean almost doubles. Interestingly, tightening credit standards does not affect the extent of adverse selection in this market. Figure 15 plotss 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.

Tightening lending standards increases the level and dispersion of equilibrium rates paid in the market through two mechanisms. First, borrowers are more willing to accept higher mortgage rates in order to avoid paying their search cost for multiple subsequent periods even holding offered rates fixed. Intuitively, borrowers behave as though they face increased search costs, which results in higher transacted prices. Second, the increased reservation rates of borrowers induces lenders to charge higher rates. Higher offered rates in turn feed back into the borrowers' reservation rates. Thus the increased probability of rejection induces price adjustment through an interaction of supply and

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

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

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

demand side effects. Overall, our model illustrates that tighter lending standards results 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. 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 Discrimination and Redlining

Redlining is a practice of discrimination that denies access to products, such as mortgages, to consumers due to their socioeconomic, racial, or ethnic makeup. The search frictions we introduce here permit the analysis of a more 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 and the interaction with adverse selection in this market.[31]

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.[32] Half of banks in the market redline: a redlining bank approves the discriminated B borrowers with a 50% smaller probability than the non-discriminated W borrowers of the same creditworthiness. The non-redlining banks ignore the B, W distinction. Lastly, banks can only discriminate based on acceptance probabilities, and have to offer the same interest rates to the discriminated and non-discriminated groups. Preventing such discrimination 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. Last, we assume that the borrowers are only aware of the proportion of banks redlining, 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 17, and summarized in Table 11.

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; discriminated borrowers sort to banks who offer higher rates. These higher rates charged by banks are not related to actual redlining decisions of banks: the rates charged by redlining and non-redlining banks are quite similar. This is because the principal determinant of a firm's

[31] Lang et al. (2014) 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.

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

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.

Becker (1957) argues that discriminating firms lose profits. In our setting, redlined lenders' profits are mostly hurt though lower volumes and higher screening costs per accepted borrower, than the actual 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. Note that both of these offered rate distributions have higher means than the 0.206pp of the baseline estimation. The intuition for this follows that of the tighter lending standards counterfactual above - the redlined borrowers increase their reservation rates, allowing banks which do not redline to 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 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, as they observe a decline in profits of 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.

8.3 Community Reinvestment Act (CRA) - Uninformative Screening

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 CRA has in fact increased credit access and the riskiness of lending (Agrawal et al. 2012, Bhutta 2012, Bostic and Breck 2003, Dahl et al. 2000). We investigate the consequences of CRA by examining the consequences of extending credit to low creditworthiness borrowers. Our model allows us to better understand the equilibrium consequences of this policy on different types of borrowers and banks. We model an extreme version of CRA, in which the probability of mortgage acceptance for both high and low type individuals is equalized. This does not remove the underlying asymmetric information, just the banks' ability to act on it.[33] This stark benchmark allows us to illustrate the largest potential consequences CRA would have in our model.

One might conjecture that if banks were prevented from screening out less creditworthy borrowers, mortgage rates would increase. Instead, CRA results in a 28.9bp decrease in average mortgage rates despite extending more credit to less creditworthy borrowers. Broadly, the introduction of CRA results in a substantial redistribution between borrowers: 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

[33] 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 9.

for high type borrowers. The equilibrium consequence of CRA is shifting credit to low creditworthiness borrowers, resulting in a large redistribution from the creditworthy borrowers. As a result, bank profits fall by 6.3bp at the mean realized interest rate, as the average creditworthiness of banks' borrowers falls. Given that \$479billion of mortgages were originated in 2017Q3,[34] this 6.3bp reduction in profits implies that screening is worth \$1.2billion $(4 * 0.00063 * \$479\text{billion})$ per year.

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; i.e. they have the same reservation rate distribution. 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.[35] Removing screening technologies has an additional effect on prices through banks' rate setting behavior. 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.

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 that consumer credit markets differ from other search markets because lenders use an approval process to evaluate 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 how consumers search for credit products, and more broadly, products in which the seller's payoff depends on buyer's characteristics, such as insurance.

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

[35] 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 16C).

References

[1] Agrawal, S., E. Benmelech, N. Bergman, and A. Seru (2012): "Did the Community Reinvestment Act (CRA) Lead to Risky Lending?," .

[2] Alexandrov, A., and S. Koulayev (2017): "No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information," .

[3] Allen, J., R. Clark, and J.-F. Houde (2013): "The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry," American Economic Review, 104(10), 3365–3396.

[4] ——— (2014): "Search frictions and market power in negotiated price markets," .

[5] Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2015): "Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market," .

[6] 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.

[7] 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.

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

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

[10] 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," .

[11] 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.

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

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

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

[15] 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.

[16] 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.

[17] Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones (2016): "Screening and Adverse Selection in Frictional Markets," .

[18] 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.

[19] Roussanov, N., H. Ruan, and Y. Wei (2017): "Marketing Mutual Funds," .

[20] 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.

10 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\} (%)	15.95	36.61	–	–
Pr\{90+ Days Delinquent\} (%)	9.65	29.53	–	–
Pr\{Prepay\} (%)	50.18	50.00	–	–
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	**Loan Data**			**Application Data**		
relative to 2006q4–2009q4:	Pre	During	Post	Pre	During	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\} (%)	18.32	18.64	1.33	–	–	–
Pr\{90+ Days Delinquent\} (%)	11.56	10.94	1.07	–	–	–
Pr\{Prepay\} (%)	46.98	54.25	48.11	–	–	–
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
Loan 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	5087.62	6462.76	8095.47	–	–	–
Investor (%)	7.22	9.05	10.93	7.09	8.19	9.84
Observations	573,891	548,819	194,097	2,609,421	1,424,922	1,324,717

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. 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^{st} Quartile	2^{nd} Quartile	3^{rd} Quartile	4^{th} Quartile
	(1)	(2)	(3)	(4)	(5)
FICO score (standardized)	-3.881***	9.256***	4.831***	-1.345***	-12.743***
	(0.091)	(0.178)	(0.289)	(0.229)	(0.332)
Combined LTV (Standardized)	0.838***	-2.853***	-0.742***	0.909***	2.686***
	(0.063)	(0.078)	(0.149)	(0.058)	(0.231)
Back-end DTI Ratio (Standardized)	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

Notes: Estimated coefficients from regression equation 1 reported. 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: Predictors of inquiry counts among realized mortgage borrowers

	# Inquiries	1^{st} Quartile	2^{nd} Quartile	3^{rd} Quartile	4^{th} Quartile
	(1)	(2)	(3)	(4)	(5)
FICO score (Standardized)	-0.389***	6.570***	0.012***	-0.007***	-7.096***
	(0.030)	(0.271)	(0.004)	(0.002)	(0.604)
Combined LTV (Standardized)	0.099***	-2.261***	0.000	0.004***	1.874***
	(0.007)	(0.129)	(0.001)	(0.001)	(0.160)
Back-end DTI Ratio (Standardized)	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. 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 5: Relationship between search and origination rates

FICO:	All	All	≤ 620	$620 - 720$	> 720
	(1)	(2)	(3)	(4)	(5)
# Inquiries	-0.191***	-0.011***	0.008***	-0.005	-0.010***
	(0.058)	(0.003)	(0.003)	(0.003)	(0.002)
# Inquiries^2	1.992***	0.165***	-0.004	0.107***	0.130***
	(0.481)	(0.034)	(0.033)	(0.029)	(0.025)
Controls	N	Y	Y	Y	Y
Observations	1293296	1280398	75251	453911	751236
R^2	0.0196	0.8040	0.5317	0.7282	0.8266
Mean of Dep. Var.	5.69	5.69	6.51	5.97	5.43
S.D. of Dep. Var.	0.86	0.86	0.71	0.77	0.82
S.D. of Inquiries	2.00	2.00	2.25	2.16	1.86

Notes: Estimated coefficients from regression equation 2 reported. Standard errors clustered at the origination quarter level reported in parentheses beneath coefficient. All regressions include lender, state, and origination quarter fixed effects. Coefficients marked with *, **, and *** are statistically different from 0 at the 10%, 5%, and 1% level, respectively.

Table 6: Linear relationship between search and origination rates within FICO groups

	FICO Score		
	< 620	620–720	720+
	(1)	(2)	(3)
Inquiries	0.008***	0.006***	0.001***
	(0.001)	(0.000)	(0.000)
Observations	75251	453911	751236
R-squared	0.4854	0.7129	0.8218
Mean Rate	6.51	5.97	5.43
S.D. Rate	0.71	0.77	0.82
S.D. Inquiries	2.25	2.16	1.86

Notes: Table reports estimated linear relationship between origination interest rate and search for our three FICO buckets. The estimated slope coefficients are pairwise statistically different at the 5% level. Standard errors clustered at the origination quarter level reported in parentheses beneath coefficient. 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 7: 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. 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 8: Relationship between search and default rates

FICO:	All	All	≤ 620	$620 - -720$	> 720
	(1)	(2)	(3)	(4)	(5)
# Inquiries	-1.265	1.006***	2.450***	2.157***	0.413**
	(0.897)	(0.300)	(0.264)	(0.264)	(0.191)
# Inquiries^2	29.071***	1.351	-7.939***	-4.448*	2.914
	(6.119)	(2.534)	(2.500)	(2.614)	(2.059)
Controls	N	Y	Y	Y	Y
Observations	1305023	1291997	82736	457657	751604
R^2	0.0061	0.2156	0.1317	0.1542	0.0961
Mean of Dep. Var.	15.68	15.63	50.33	25.03	6.08
S.D. of Dep. Var.	36.37	36.31	50.00	43.32	23.90
S.D. of Inquiries	2.00	2.00	2.29	2.17	1.86

Notes: Estimated coefficients from regression equation 2 reported. Standard errors clustered at the origination quarter level reported in parentheses beneath coefficient. All regressions include lender, state, and origination quarter fixed effects. Coefficients marked with *, **, and *** are statistically different from 0 at the 10%, 5%, and 1% level, respectively.

Table 9: 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 13.3.

Table 10: 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
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
Uninformative Screening - CRA	0.007	0.463	0.005	0.462	0.008	0.463
Uninformative Screening - CRA (Eqm.)	-0.291	0.471	-0.293	0.471	-0.290	0.471
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
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
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
Uninformative Screening - CRA	2.76	2.08	2.76	2.08	2.76	2.07
Uninformative Screening - CRA (Eqm.)	2.78	2.09	2.78	2.09	2.77	2.09
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

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
Tighter Lending Standards	0.142	0.547	1.860
Tighter Lending Standards (Eqm.)	0.483	0.805	2.130
Uninformative Screening - CRA	0.142	1.873	1.867
Uninformative Screening - CRA (Eqm.)	-0.147	0.561	1.830
Redlining	0.142	0.547	1.860
Redlining (Eqm.)	0.300	0.733	1.995

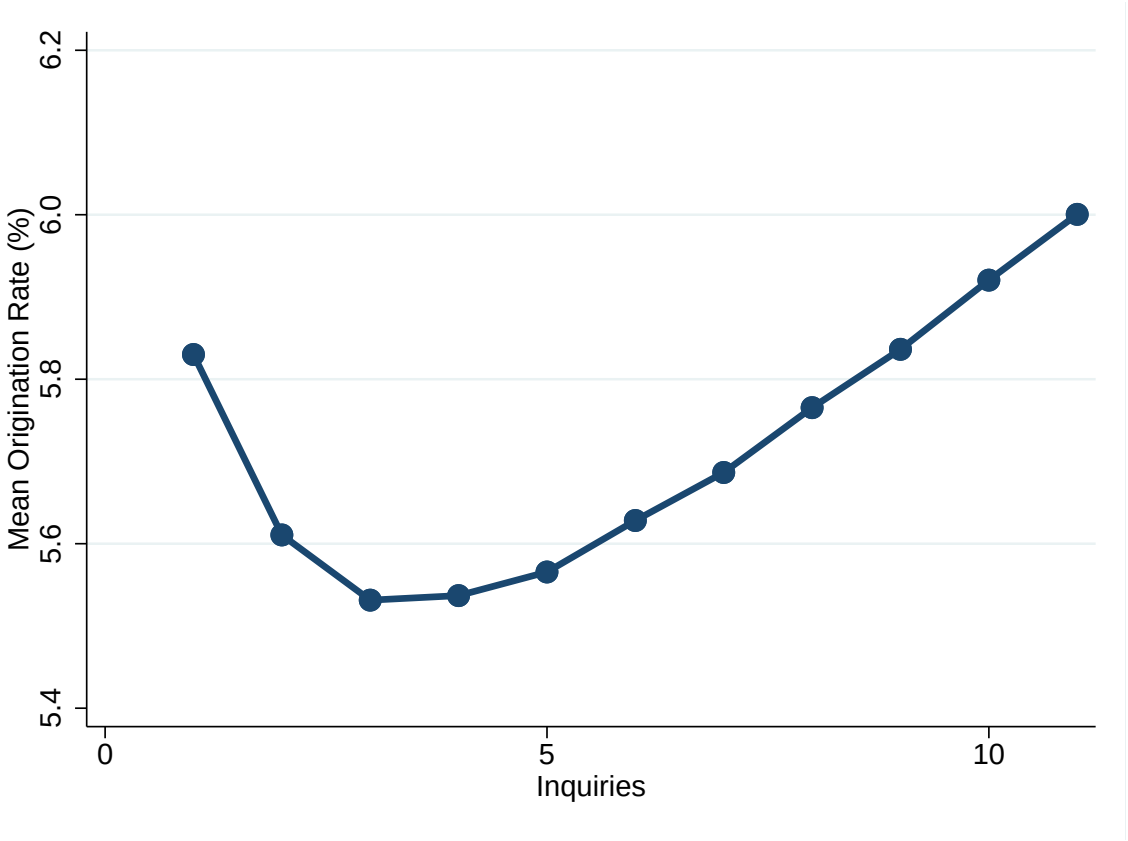
Notes: Table reports mean and standard deviation of search and realized interest rates across our counterfactual model simulations. The first two columns report mean and standard deviations for the full simulated sample of borrowers. The third and fourth columns report the mean and standard deviation 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}.

Table 11: 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	
	Average	S.D.	Profits	
Redlining Banks	0.291	0.753	1.867	
Non-Redlining Banks	0.308	0.713	2.124	

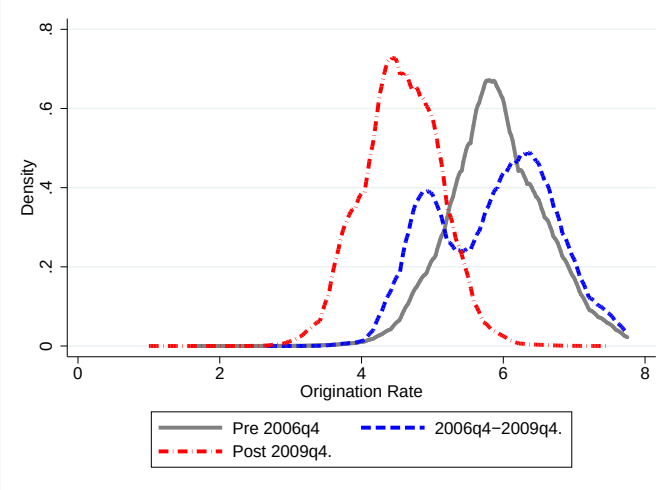
Notes: Table reports mean and standard deviation of search and realized interest rates for the group of redlined borrowers B and borrowers not subject to redlining, W. The first panel reports the effect of redlining on borrowers, while the second panel reports the effect of redlining on banks. The first 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 Section 14.2. 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: Search and Mortgage Rates Paid

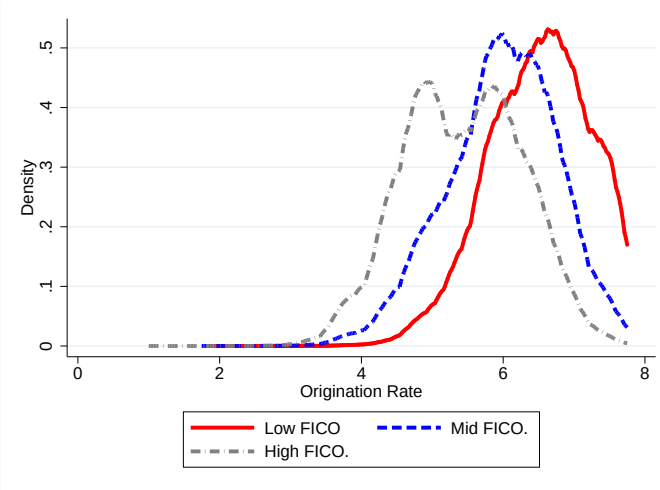


Notes: Figure plots average realized interest rates against inquiry counts for realized loans.

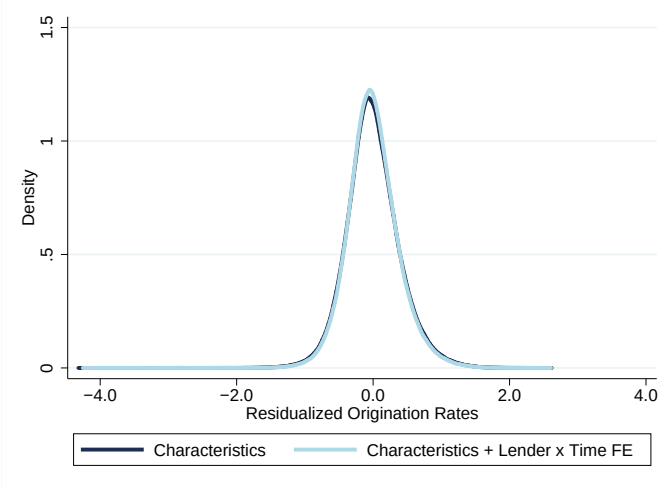
Figure 2: Distribution of Mortgage Rates in the U.S.



Panel A: Raw Rates by Origination Date



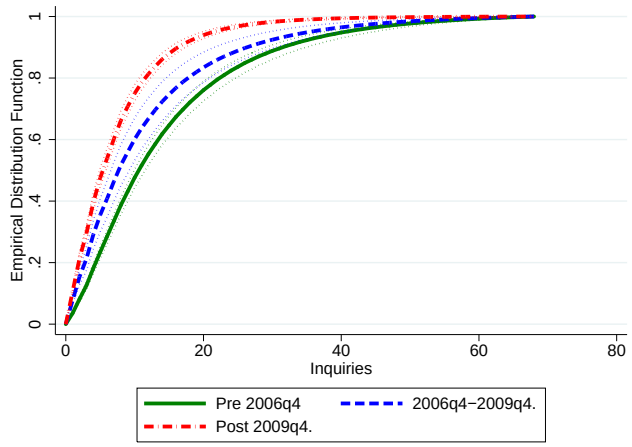
Panel B: Raw Rates by Borrower FICO Score



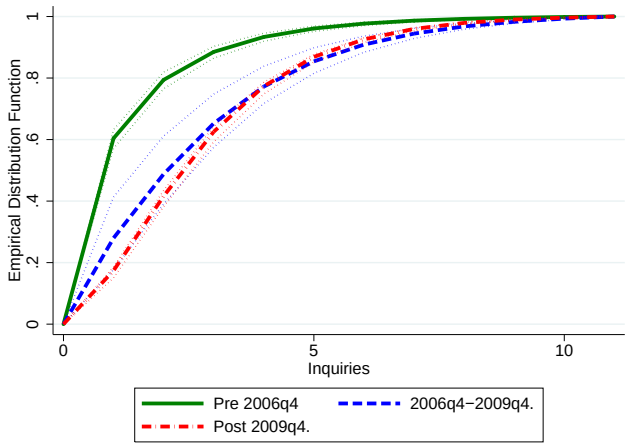
Panel C: Rates Residualized Against Observables

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.

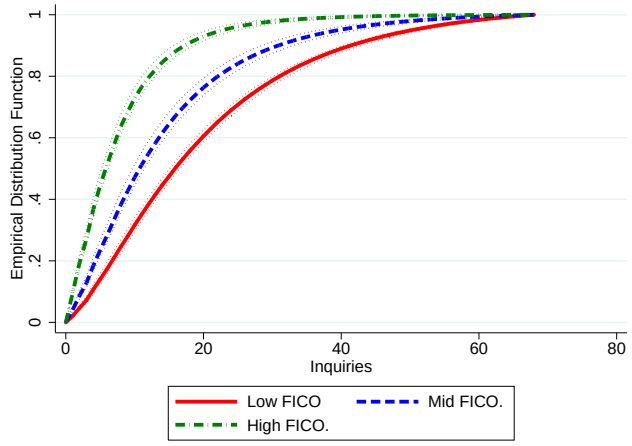
Figure 3: Inquiry distribution among mortgage applicants



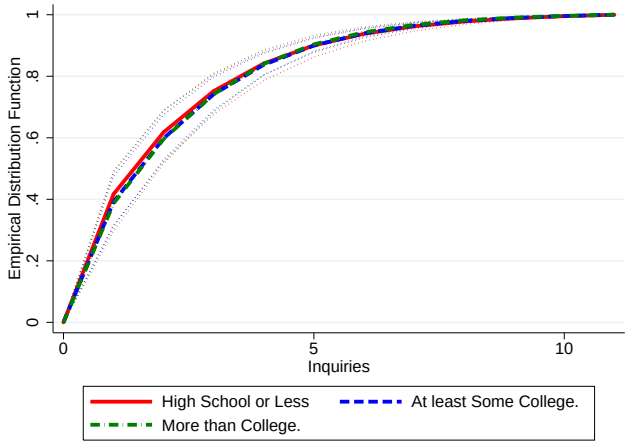
Panel A: Applicant Dataset



Panel B: Loan-Level Dataset



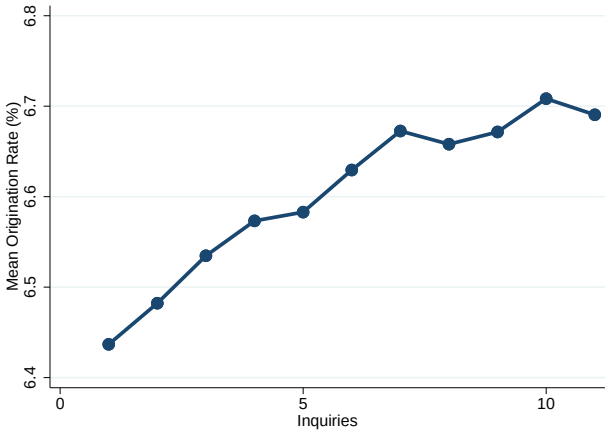
Panel C: Applicants by FICO Score



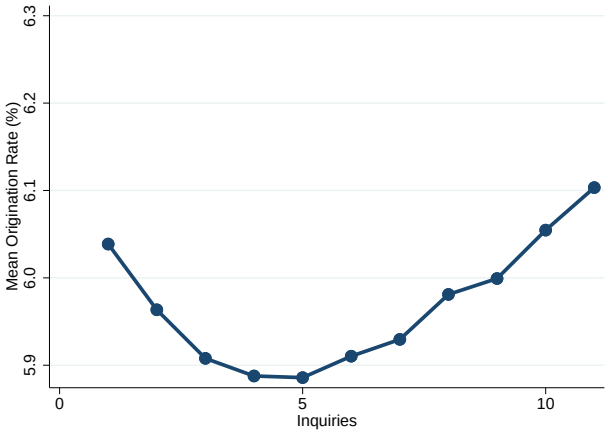
Panel D: Applicants by Education

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.

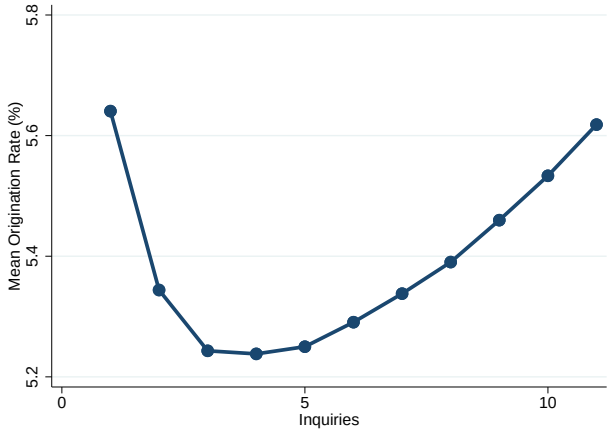
Figure 4: Rates and search by FICO bucket



Panel A: FICO $\leq$620



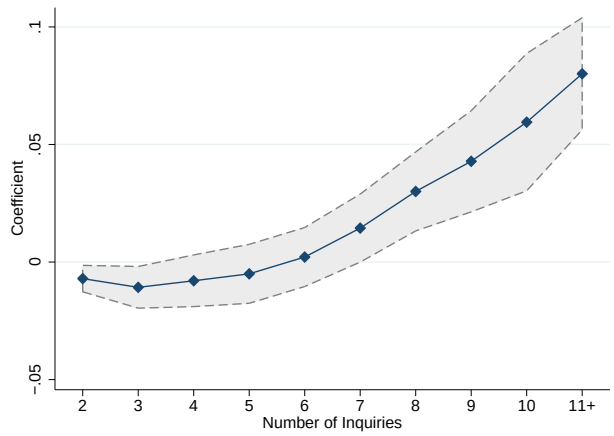
Panel B: 620 $<$ FICO $\leq$ 720



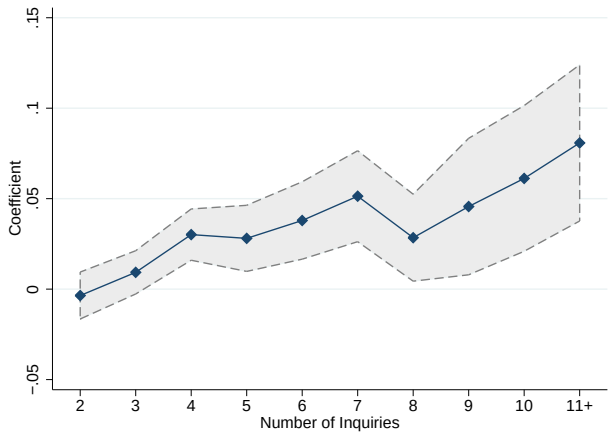
Panel C: FICO $>$ 720

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

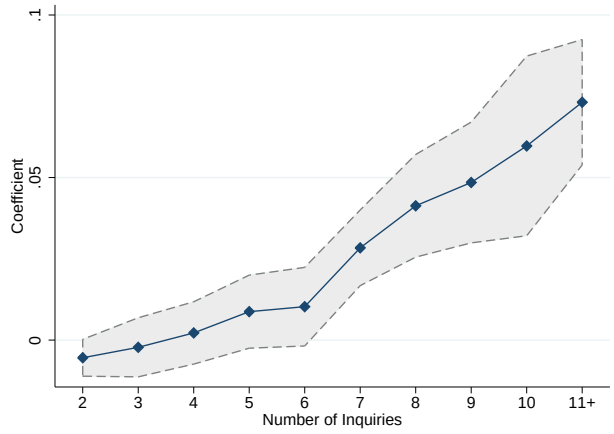
Figure 5: Relationship Between Search and Mortgage Origination Rates, Conditional on Observables



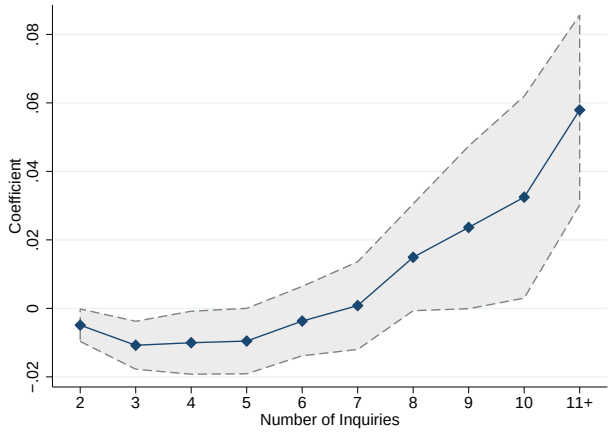
Panel A: All Borrowers



Panel B: FICO ≤620



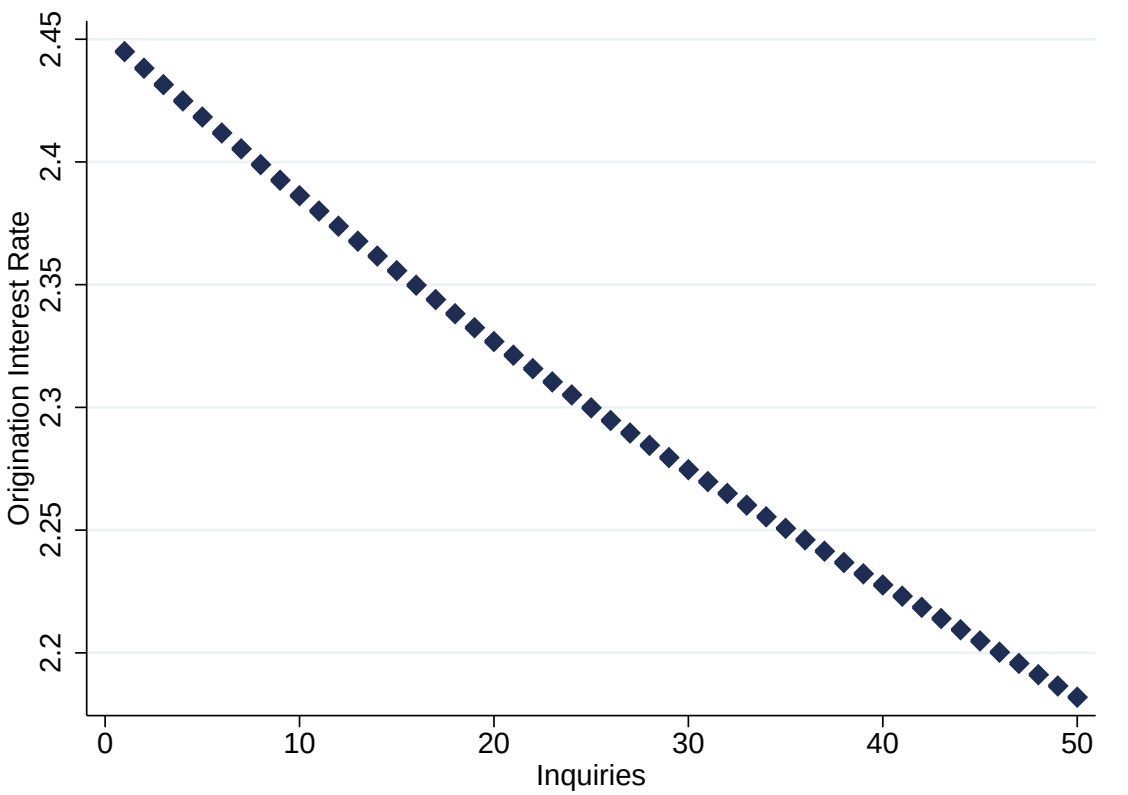
Panel C: 620 < FICO ≤ 720



Panel D: FICO > 720

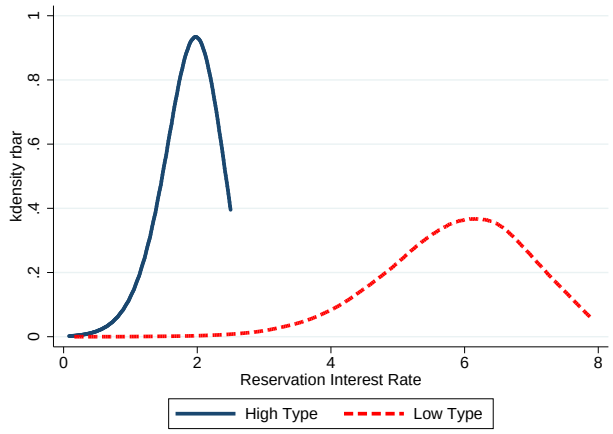
Notes: Figure plots regression coefficients estimated from equation 2 using OLS across three FICO sub-samples. The dependent variable in each regression is the origination interest rate on a loan. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination 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 6: Relationship between search and prices in standard search model without screening

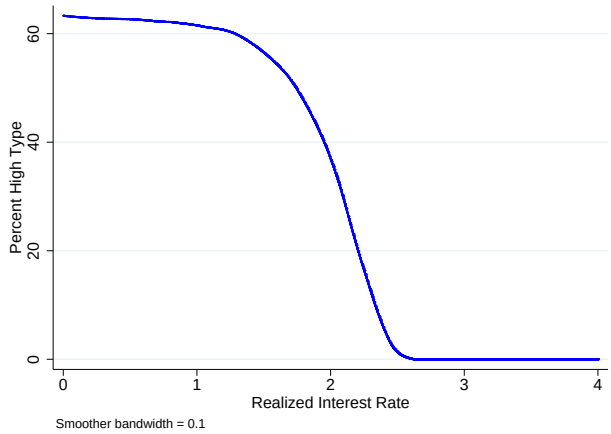


Notes: Figure plots the relationship between origination interest rates and search in the absence of screening: where $p_h = p_l = 1$, and the search costs and offered rates are distributed according to truncated normal distributions.

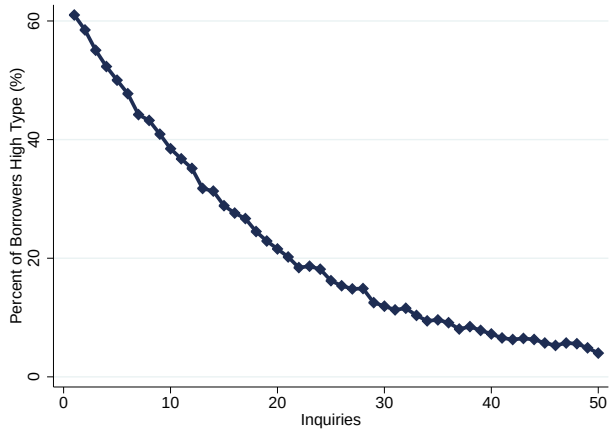
Figure 7: Characteristics of a Sequential Search Model with Informative Screening



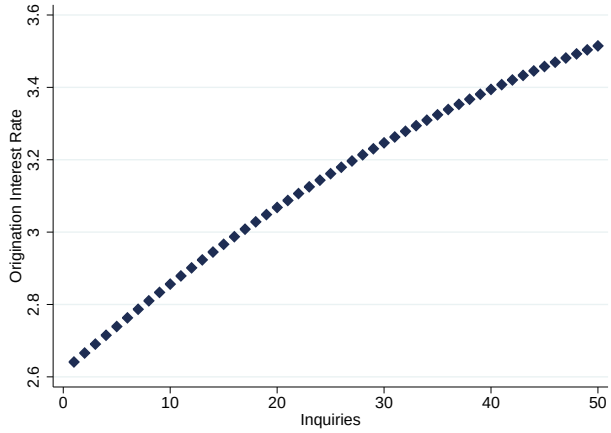
Panel A: Reservation rates of high and low types



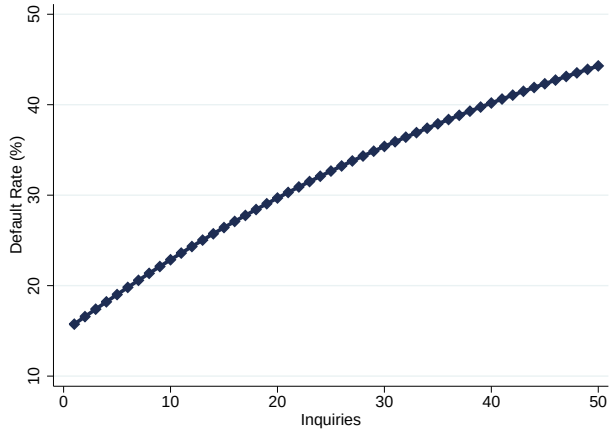
Panel B: Share of high types as function of origination rate



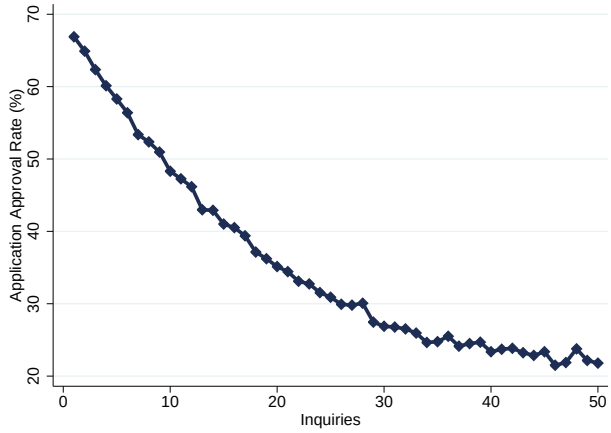
Panel C: Borrower type distribution and search



Panel D: Relationship between search and prices



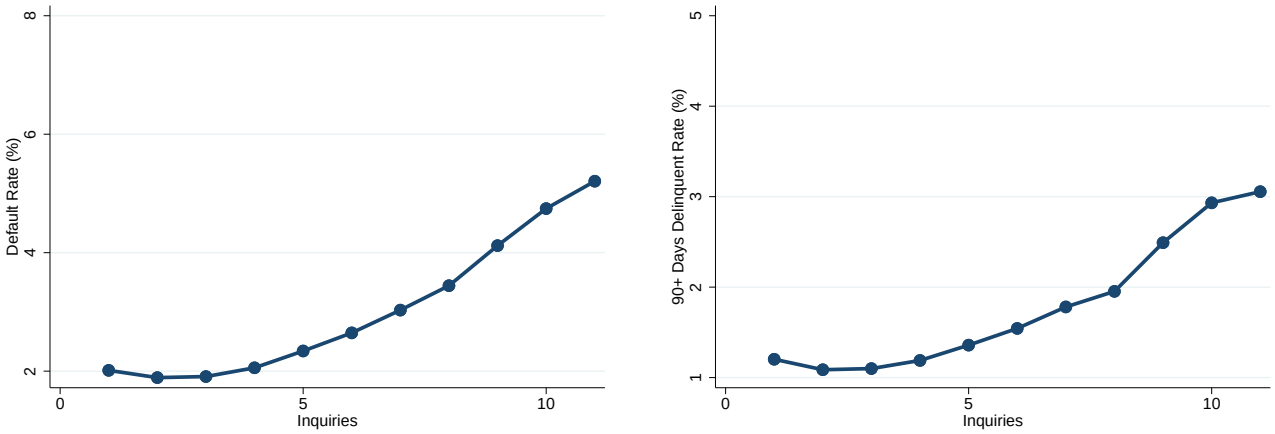
Panel E: Relationship between search and default rate



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 8: Search and Annualized Default Rate

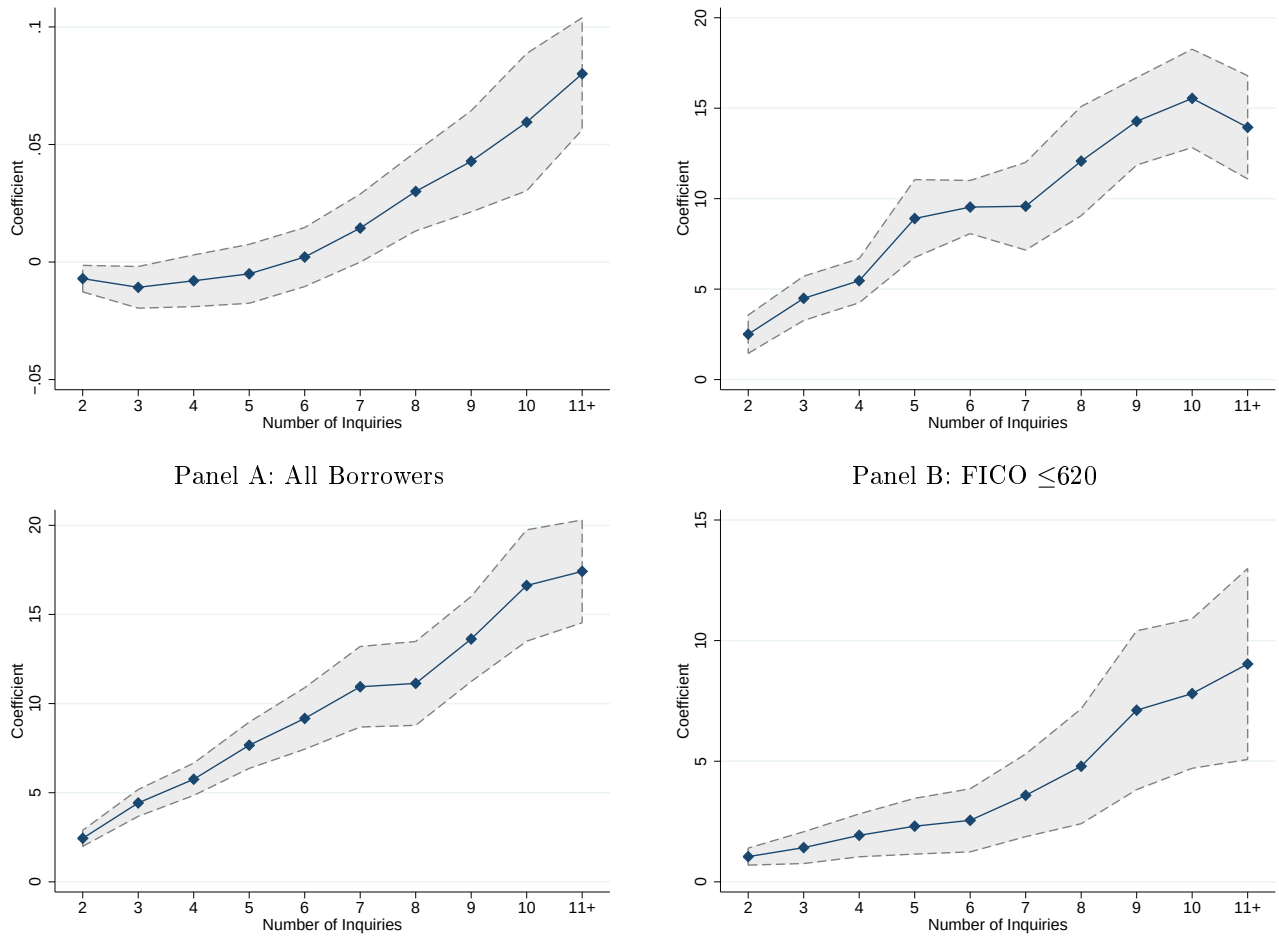


Panel A: Default

Panel B: 90+ Days Delinquent

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.

Figure 9: Relationship Between Search and Mortgage Default Rates, Conditional on Observables



Panel A: All Borrowers

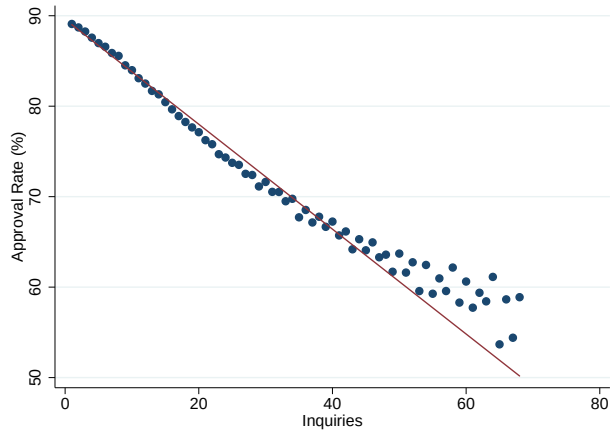
Panel B: FICO $\leq$620

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

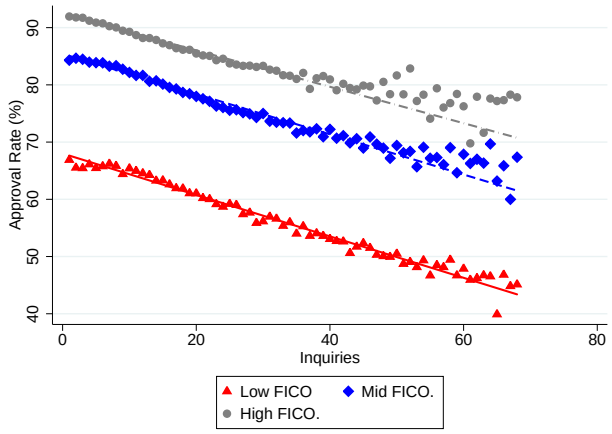
Panel D: FICO $>$ 720

Notes: Figure plots regression coefficients estimated from equation 7 using OLS. The dependent variable in each regression is an indicator for whether a loan had defaulted as of January 2015, scaled by 100 for legibility. 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 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.

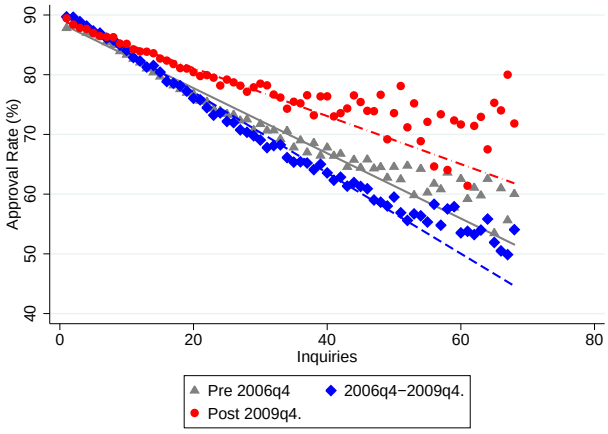
Figure 10: Relationship Between Search and Mortgage Application Approval Rates



Panel A: All Applicants



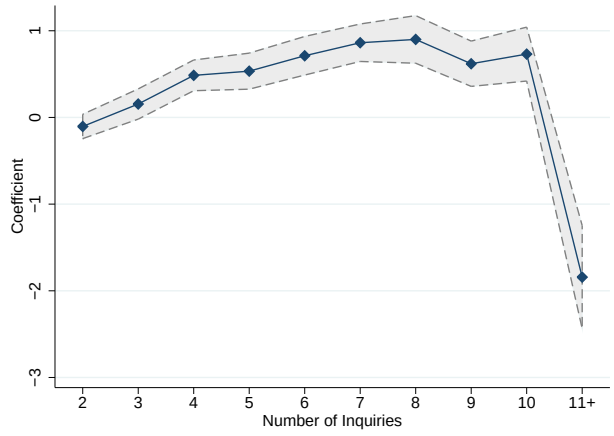
Panel B: By Borrower FICO Score



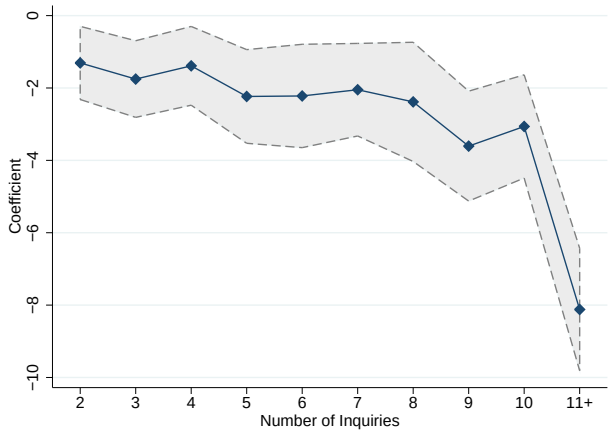
Panel C: By Origination Date

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.

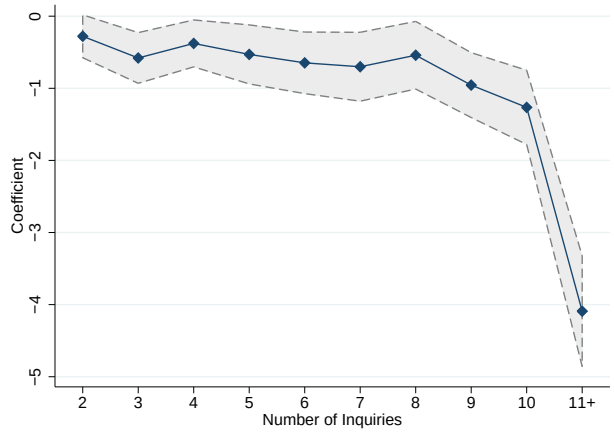
Figure 11: Relationship between search and mortgage application approval rates, conditional on observables by FICO bucket



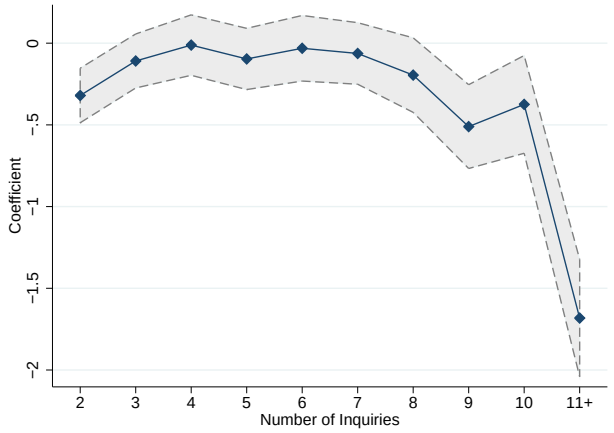
Panel A: All Applicants



Panel B: FICO ≤620



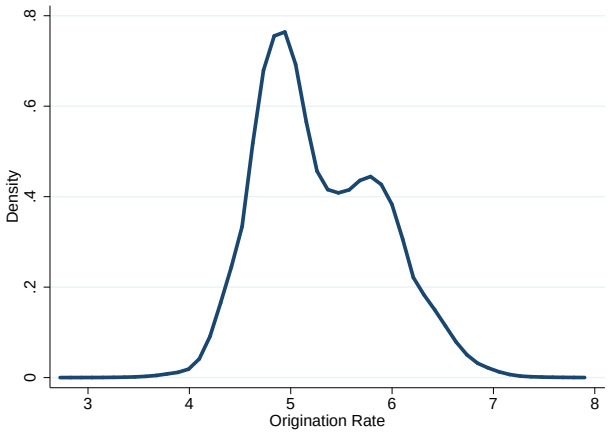
Panel C: 620 < FICO ≤ 720



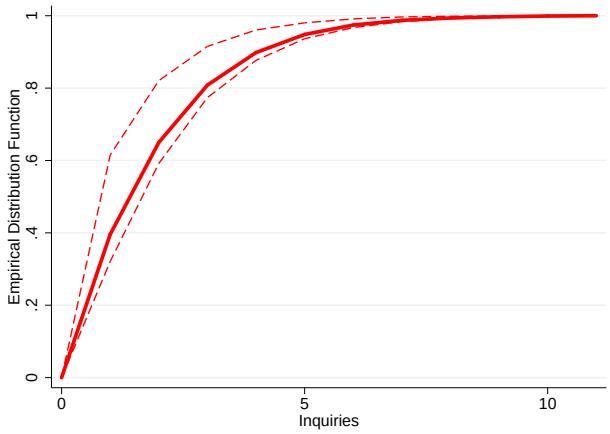
Panel D: FICO > 720

Notes: Figure plots regression coefficients estimated from equation 8 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.

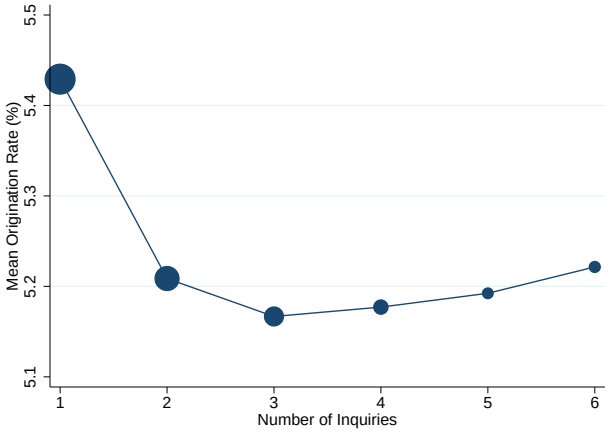
Figure 12: Search Behavior of Rarely Rejected Borrowers



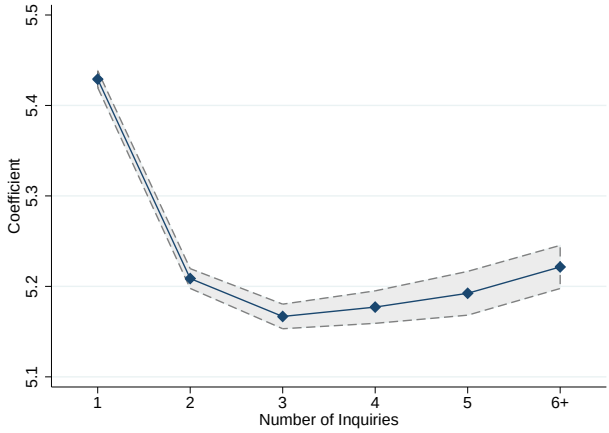
Panel A: Realized rate distribution



Panel B: Search distribution



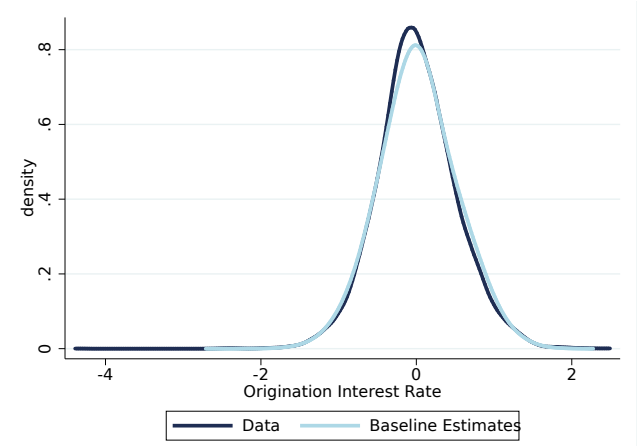
Panel C: Relationship between search and prices



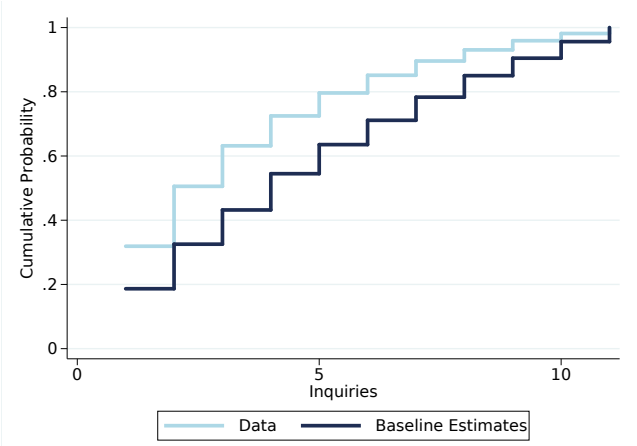
Panel D: Relationship between search and prices controlling for borrower observables

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 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.

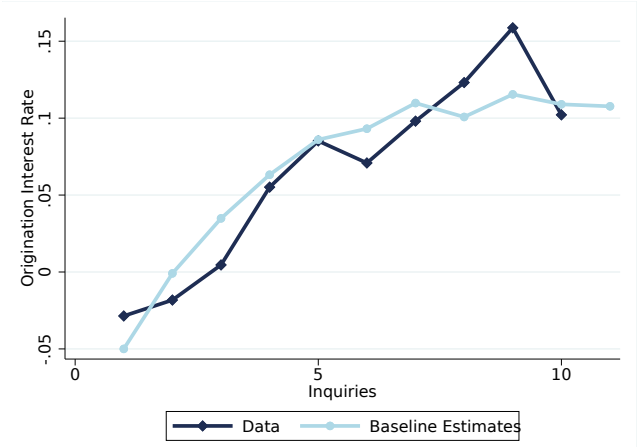
Figure 13: Model Performance: Search Behavior in Data Versus Model Simulation with Estimated Parameters



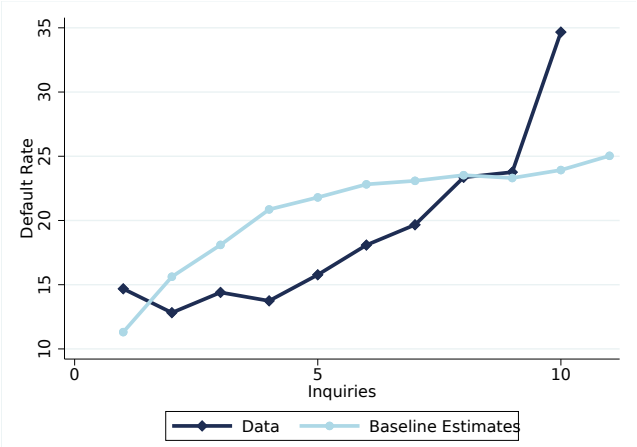
Panel A: Realized Residual Rate Densities



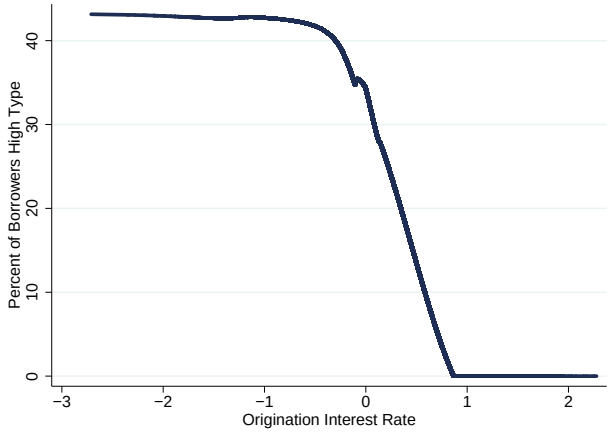
Panel B: Search Distribution



Panel C: Search and Origination Rates



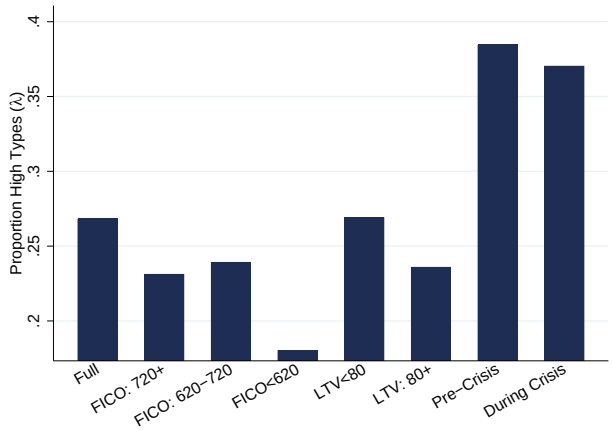
Panel D: Search and Default Probability



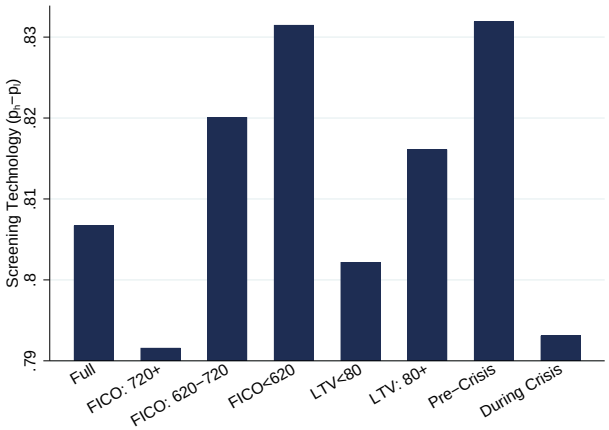
Panel E: Share of high types as function of origination rate

Notes: Figure plots the performance of our model under our benchmark estimated parameters from Table 9. 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[52]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.

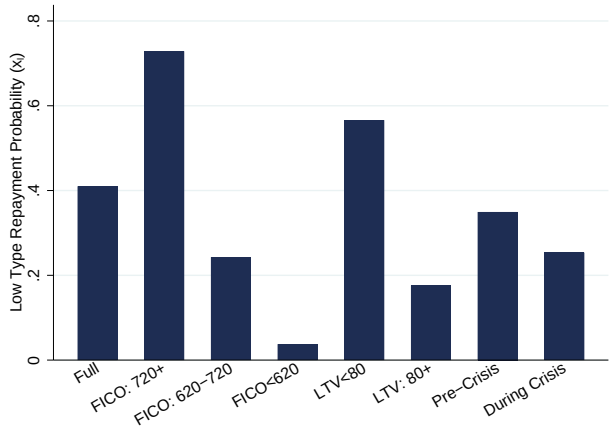
Figure 14: Estimates by subsample



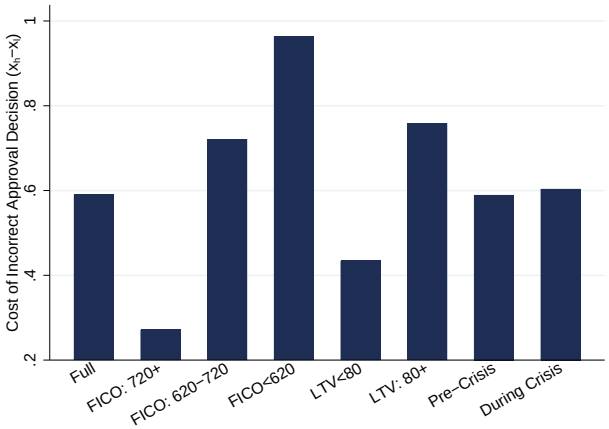
Panel A: Proportion of high types λ



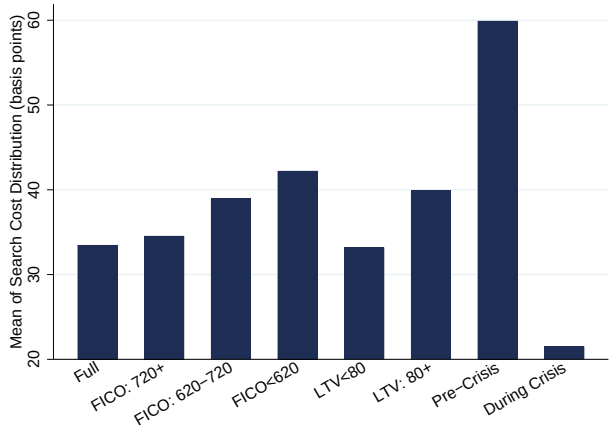
Panel B: Screening technology power $p_h - p_l$



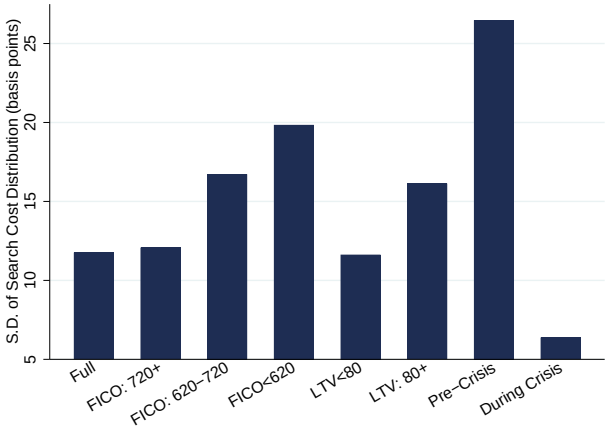
Panel C: Repayment probability of low types x_l



Panel D: Cost of misclassification $x_h - x_l$



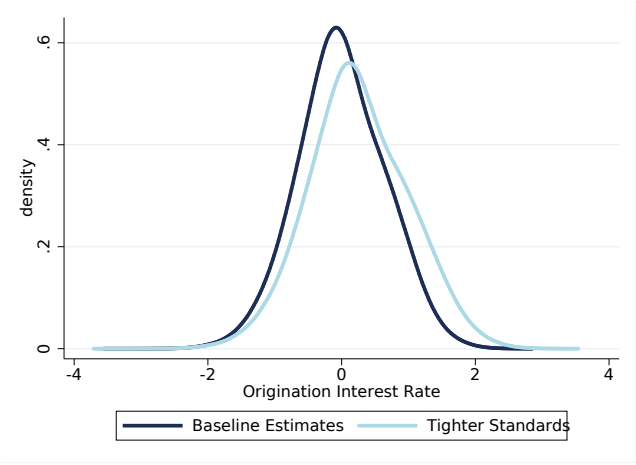
Panel E: Mean search cost $e^{(\mu_c+\sigma_c^2/2)}$



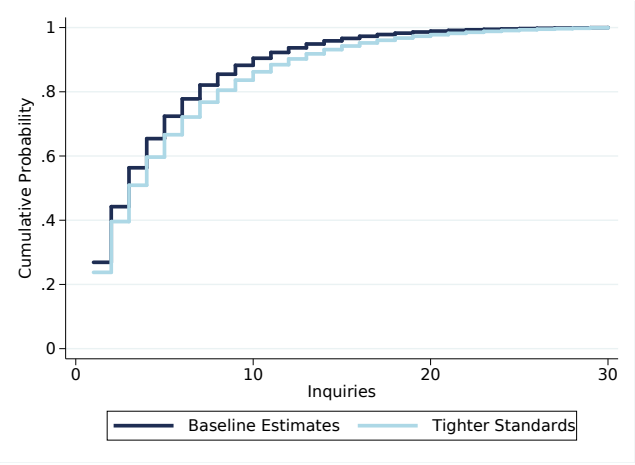
Panel F: Standard deviation of search cost $\sqrt{\left(e^{\sigma_c^2}-1\right)e^{(2\mu_c+\sigma_c^2)}}$

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.

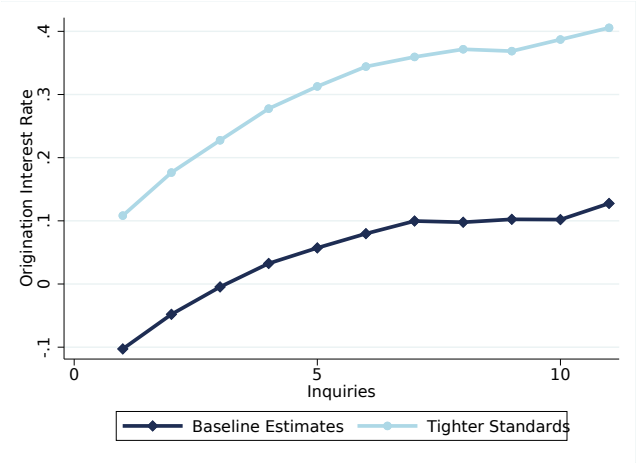
Figure 15: Tighter Lending Standards Counterfactual



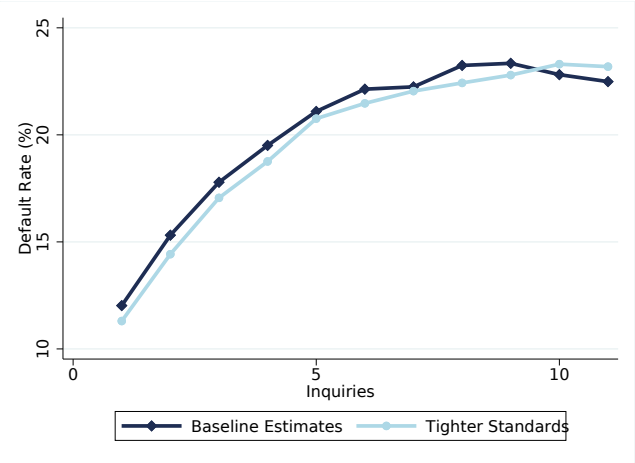
Panel A: Realized Rate Distribution



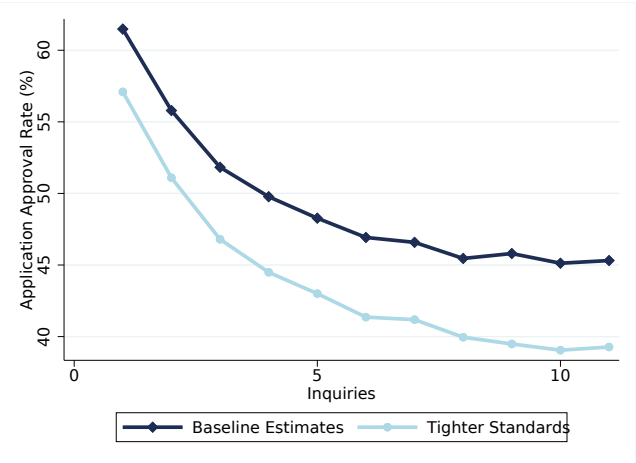
Panel B: Distribution of Search



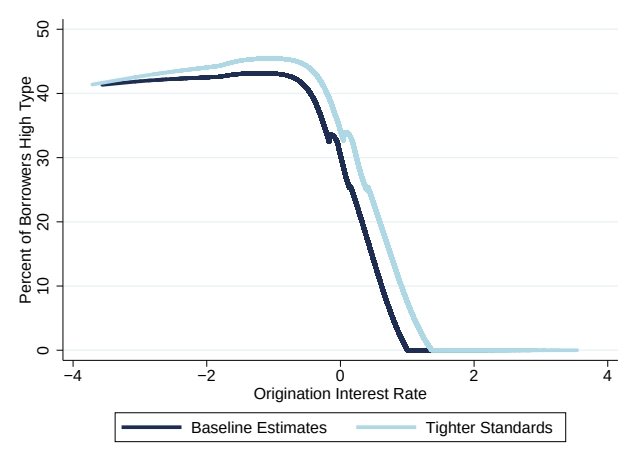
Panel C: Relationship between search and price



Panel D: Relationship between search and default



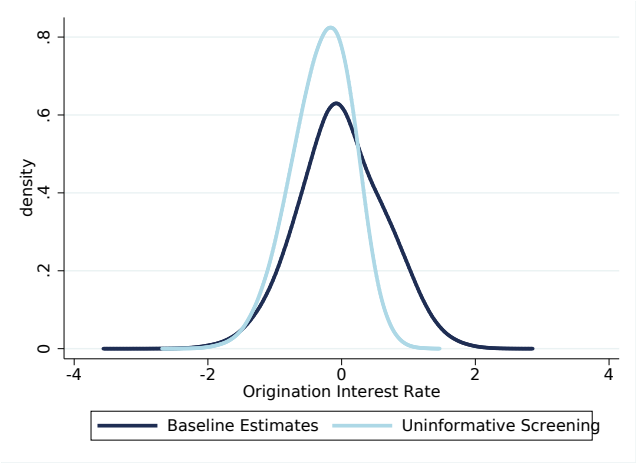
Panel E: Relationship between search and application approval



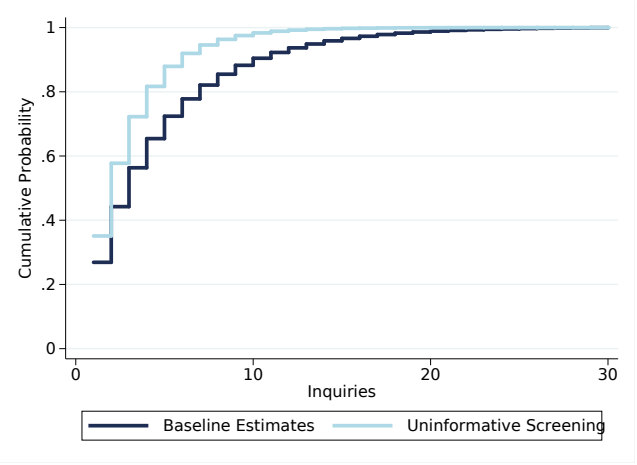
Panel F: Share of high types as function of origination rate

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. 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.

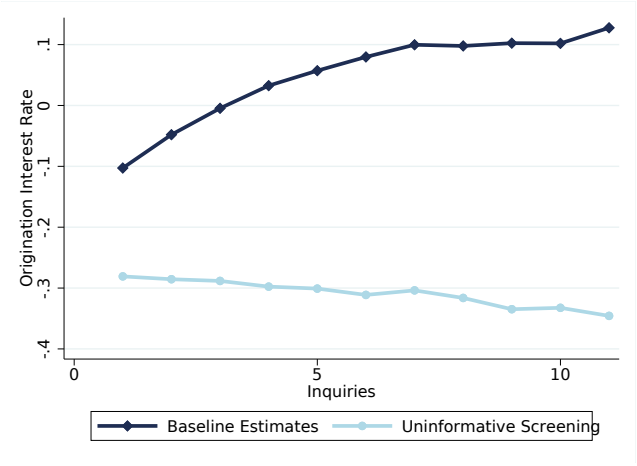
Figure 16: Uninformative Screening Counterfactual



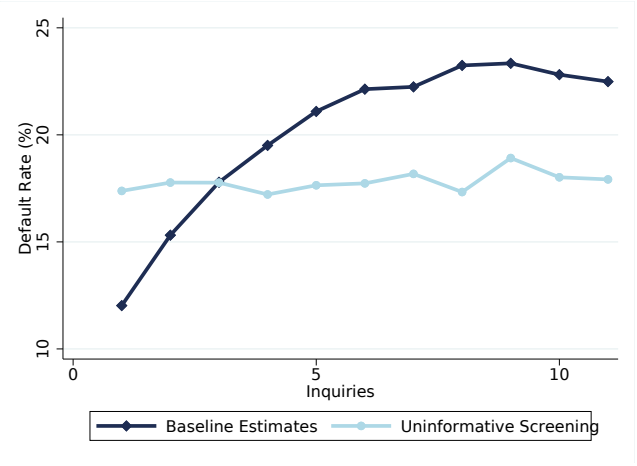
Panel A: Realized Rate Distribution



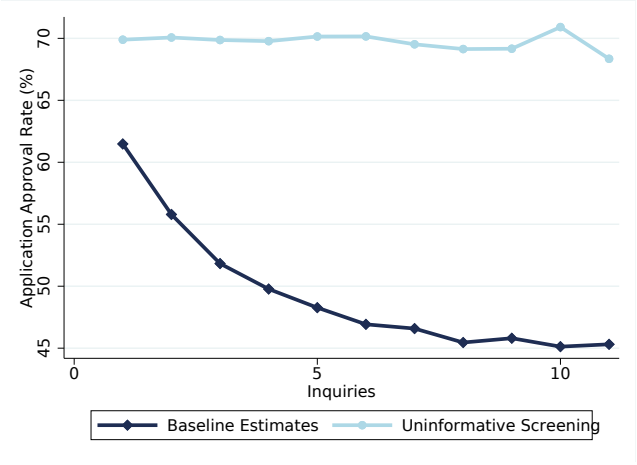
Panel B: Distribution of Search



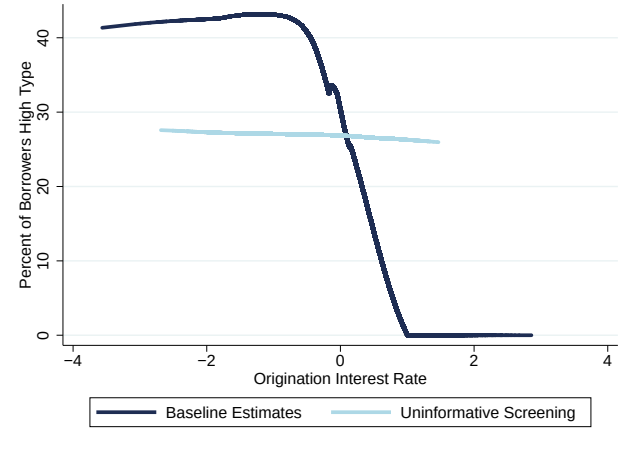
Panel C: Relationship between search and price



Panel D: Relationship between search and default



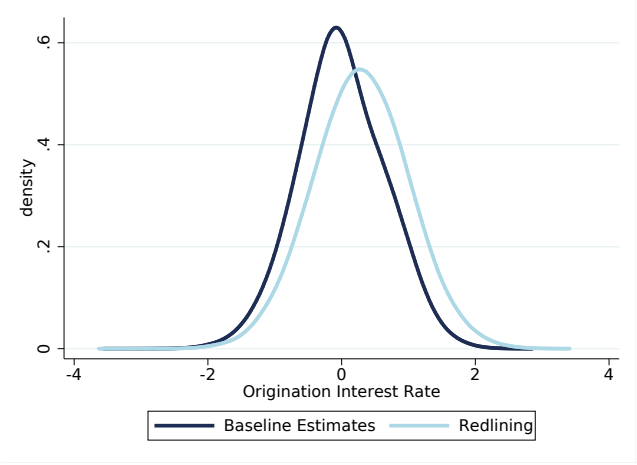
Panel E: Relationship between search and application approval



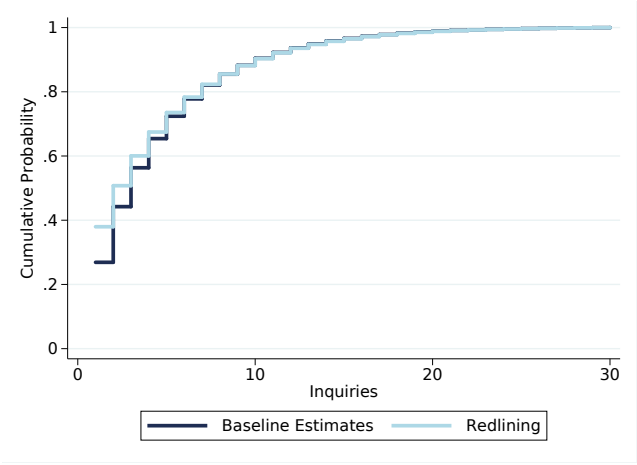
Panel F: Share of high types as function of origination rate

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.

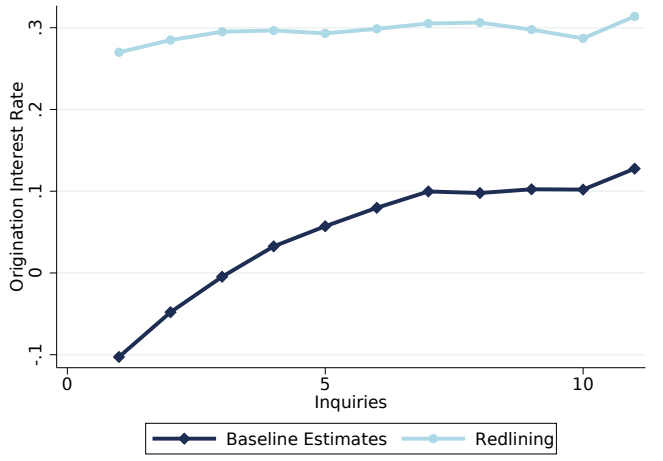
Figure 17: Redlining Counterfactual



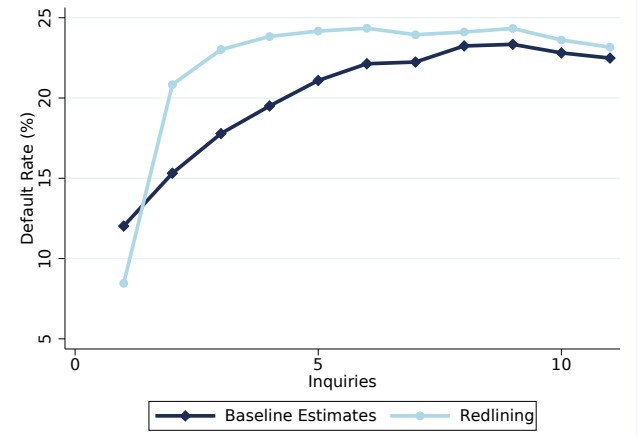
Panel A: Realized Rate Distribution



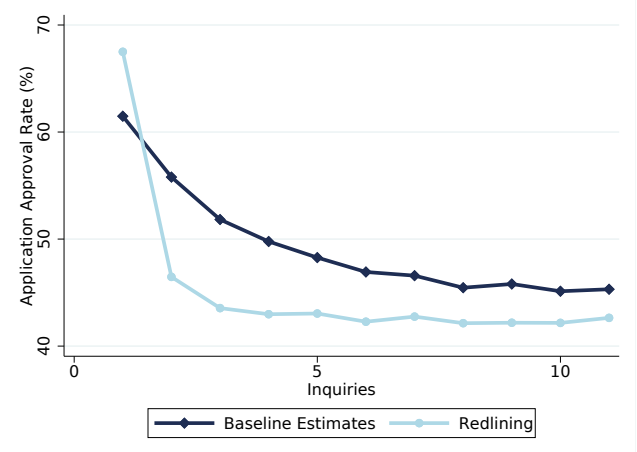
Panel B: Distribution of Search



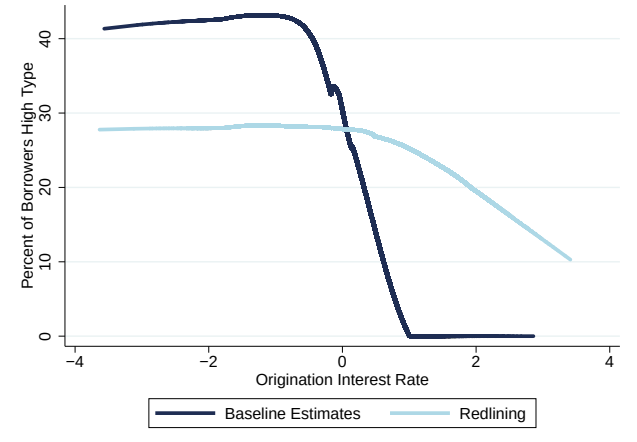
Panel C: Relationship between search and price



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 a model of redlining in which half of the lenders accept both high and low type borrowers at half of the prevailing rate. Equilibrium rates 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.

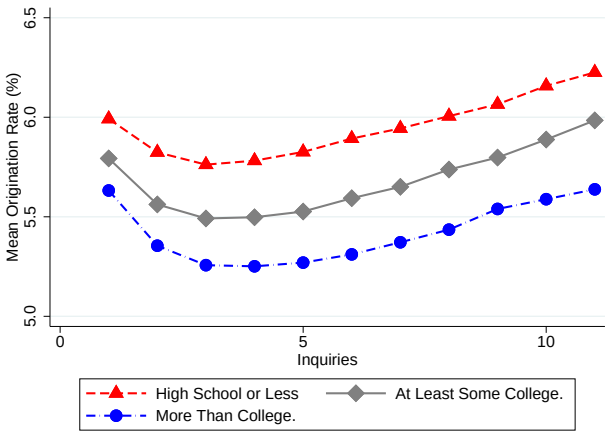
11 Appendix A: Additional Robustness Tables and Figures

Table 12: k-means Clustering Test for Multiple Borrower Types

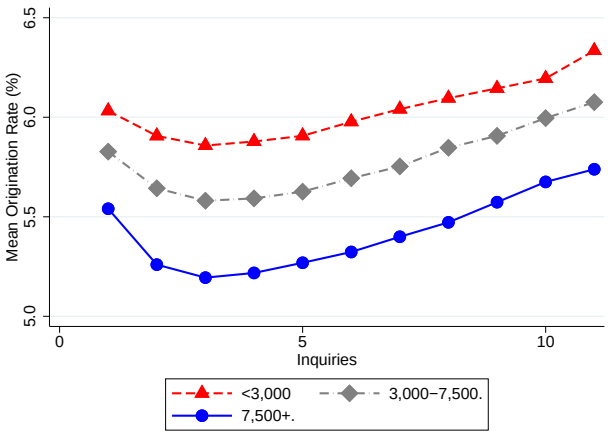
	OLS Residuals		Logit Residuals	
	(1)	(2)	(3)	(4)
	2 types	3 types	2 types	3 types
Default				
Type 1	0.000	0.000	0.000	0.000
	(1,106,792)	(474,524)	(1,106,792)	(406,650)
Type 2	1.000	0.000	1.000	0.000
	(210,015)	(632,268)	(210,015)	(700,142)
Type 3	-	1.000	-	1.000
	-	(210,015)	-	(406,650)
Approval				
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.

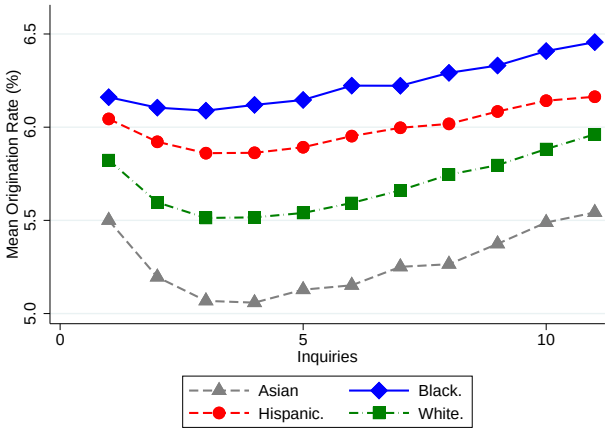
Figure 18: Mean Rates and Search by Borrower Observables



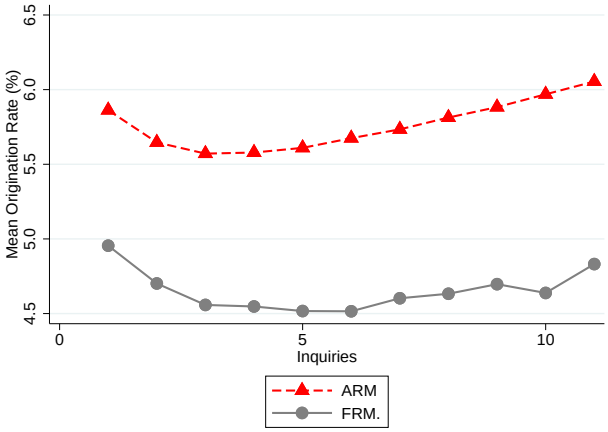
Panel A: By Education



Panel B: By Monthly Income

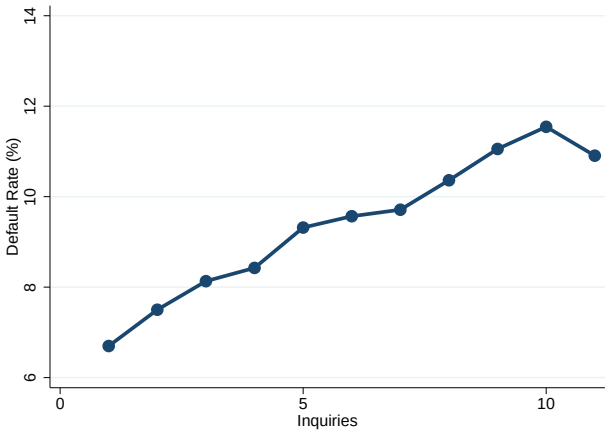


Panel C: By Race

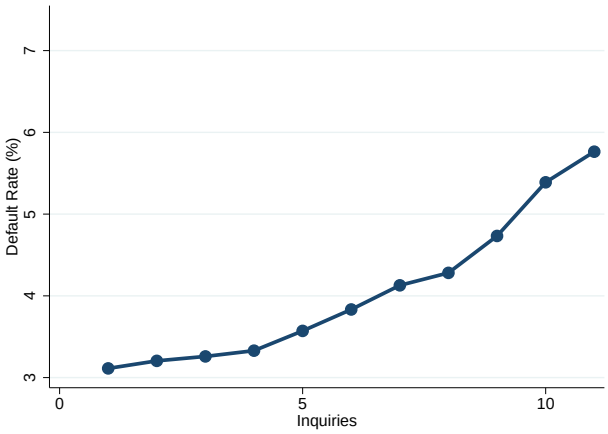


Panel D: By Product Type

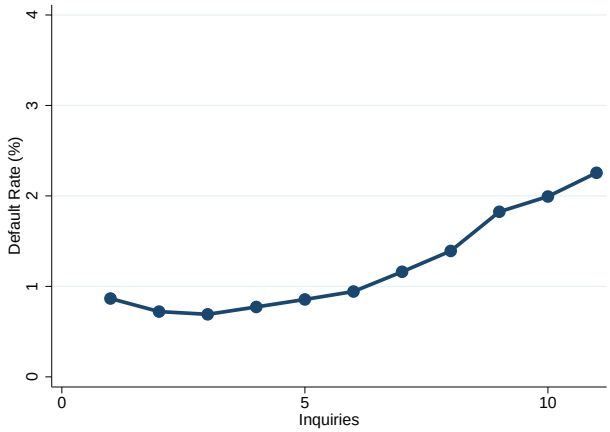
Figure 19: Default rate and search: by FICO bucket



Panel A: FICO $\leq$620

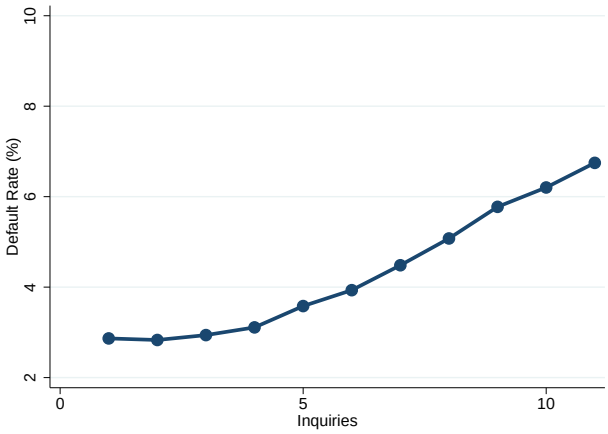


Panel B: 620 $<$ FICO $\leq$ 720

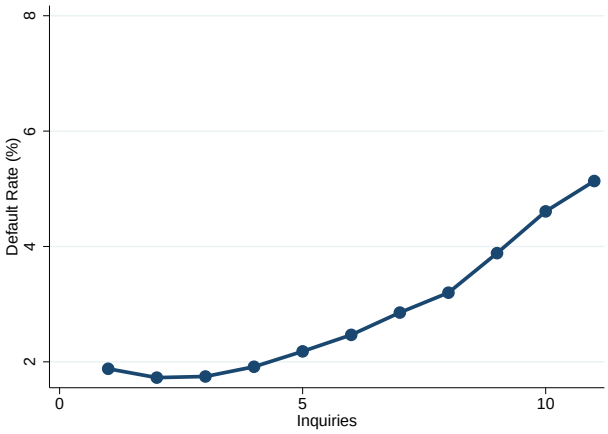


Panel C: FICO $>$ 720

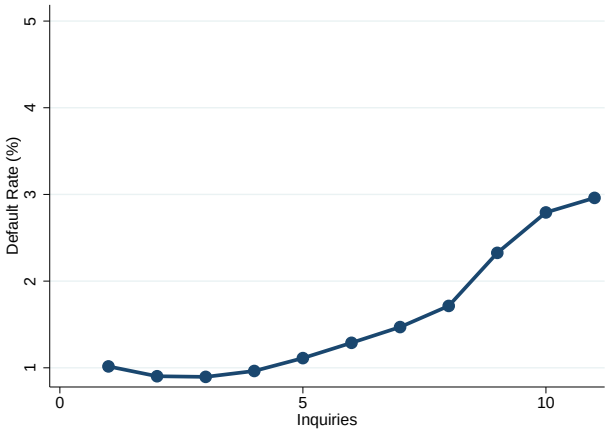
Figure 20: Default rate and search: by education level



Panel A: High School or Less

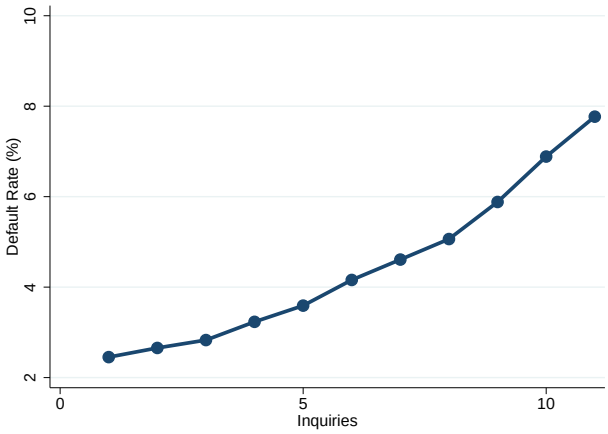


Panel B: Some College

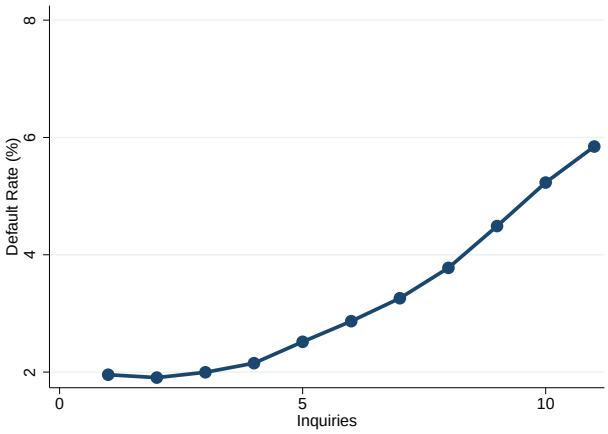


Panel C: College Graduate

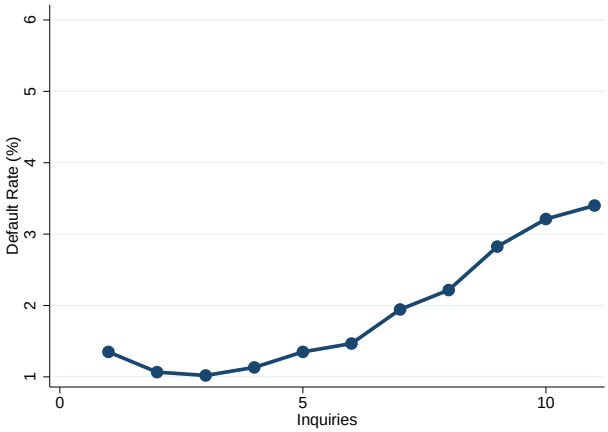
Figure 21: Default rate and search: by monthly income group



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

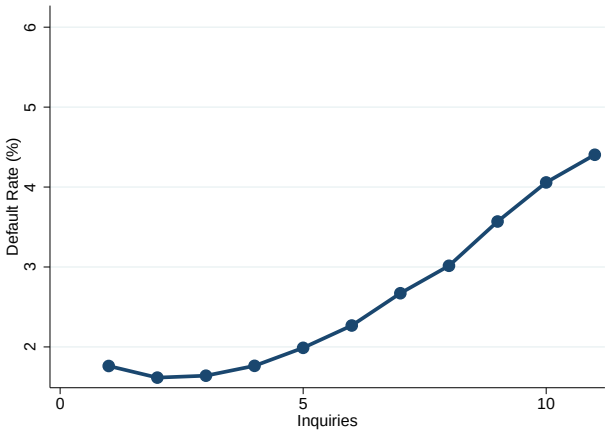


Panel B: \$3,000 $<$ Income$\leq$ \$7,500

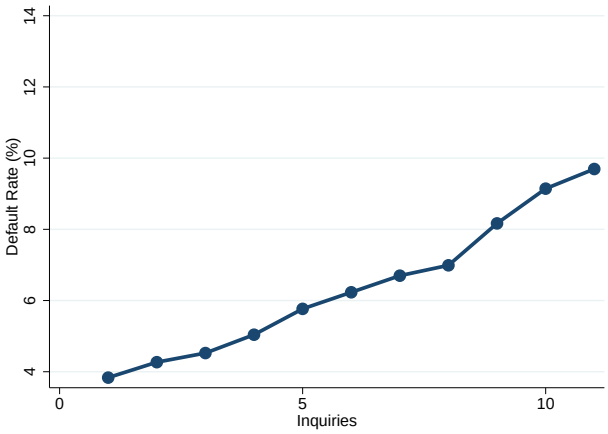


Panel C: Monthly Income $>$ \$7,500

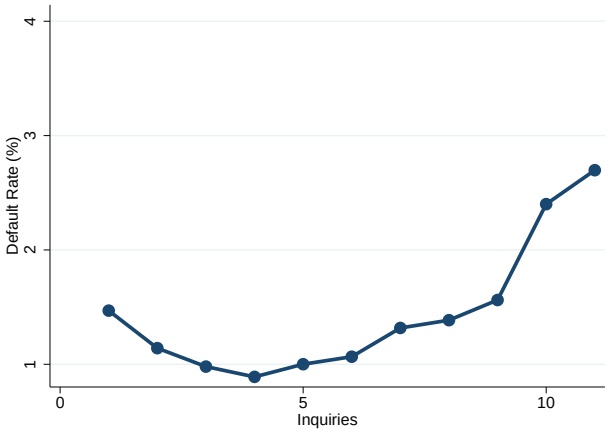
Figure 22: Default rate and search: by borrower race



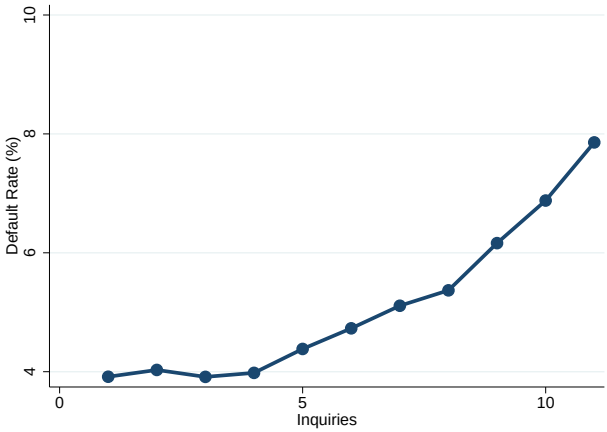
Panel A: White



Panel B: Black

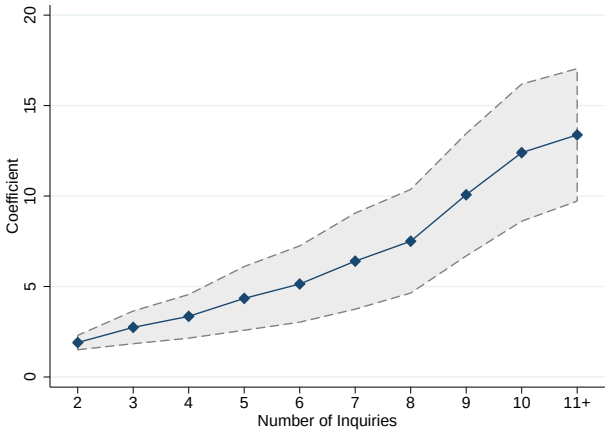


Panel C: Asian

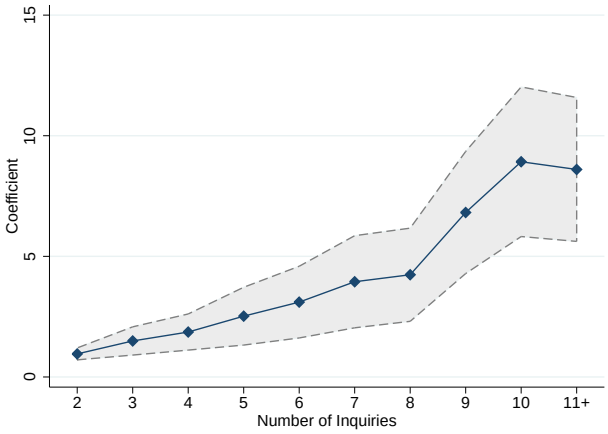


Panel D: Hispanic

Figure 23: Coefficient Plots from Regression of Default Rates on Inquiry Counts and Controls

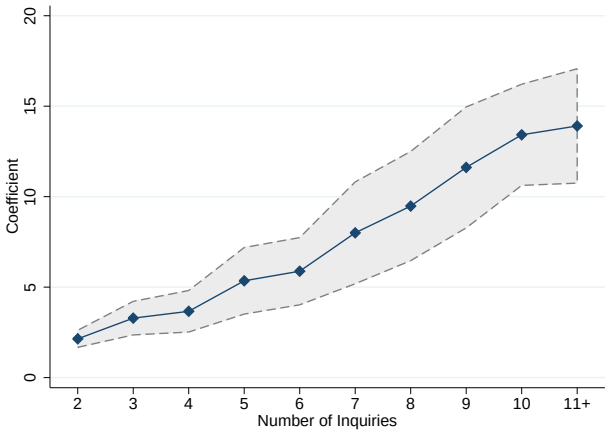


Panel A: Default Indicator

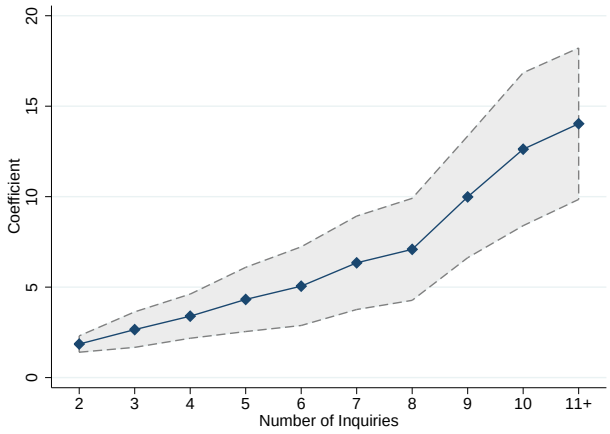


Panel B: 90+ Days Delinquent Indicator

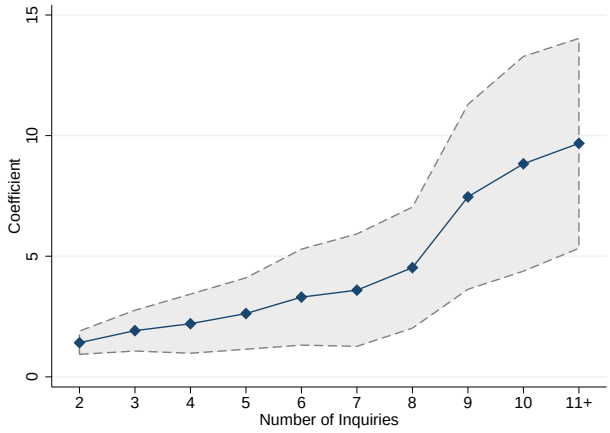
Figure 24: Coefficient Plots from Default Regression 2 - by Education Group Subsample



Panel A: High School or Less

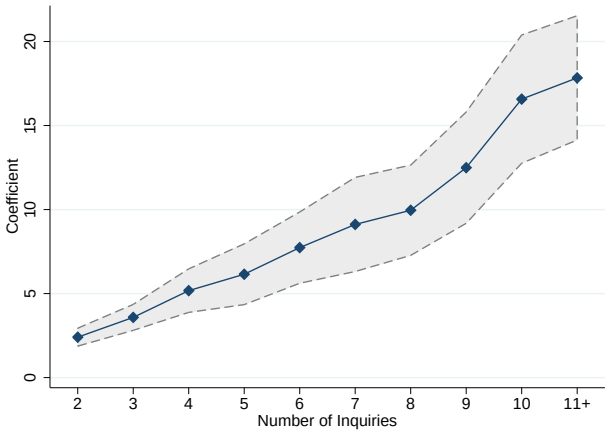


Panel B: Some College

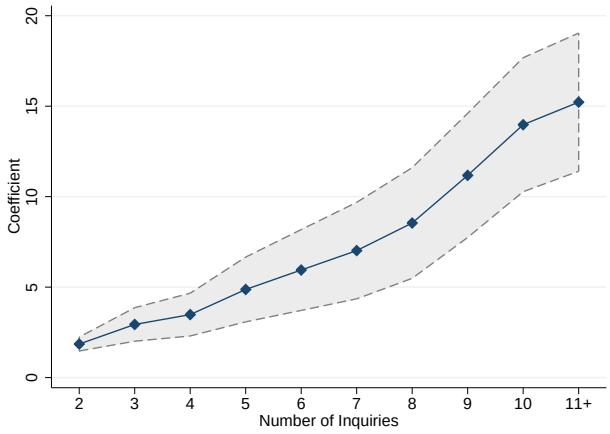


Panel C: College Graduate

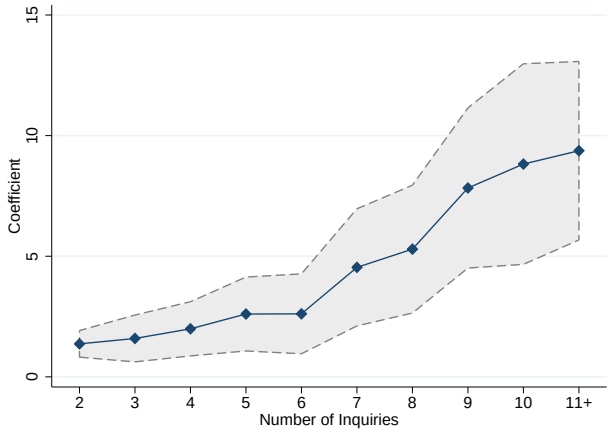
Figure 25: Coefficient Plots from Default Regression 2 - by Income Group Subsample



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

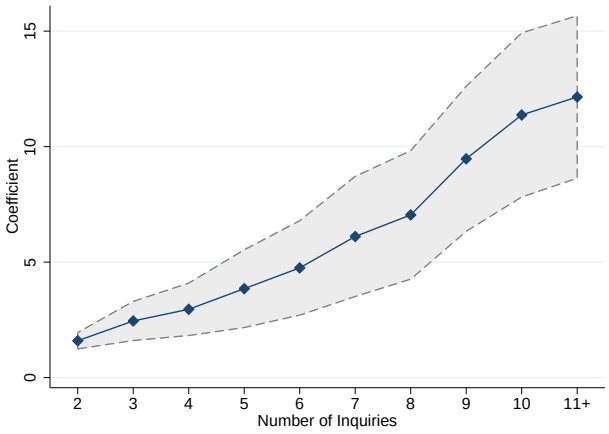


Panel B: \$3,000 $<$ Monthly Income $\leq$\$7,500

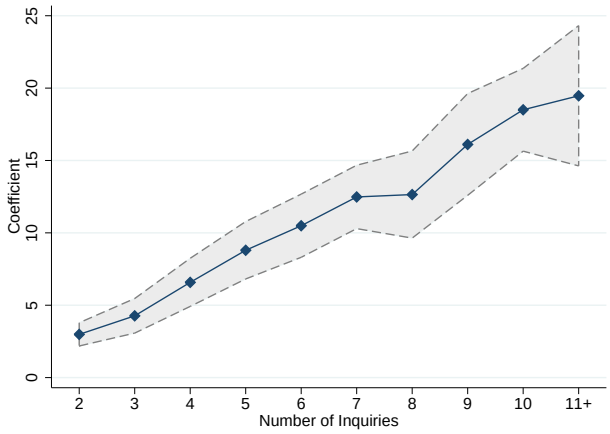


Panel C: Monthly Income $>$ \$7,500

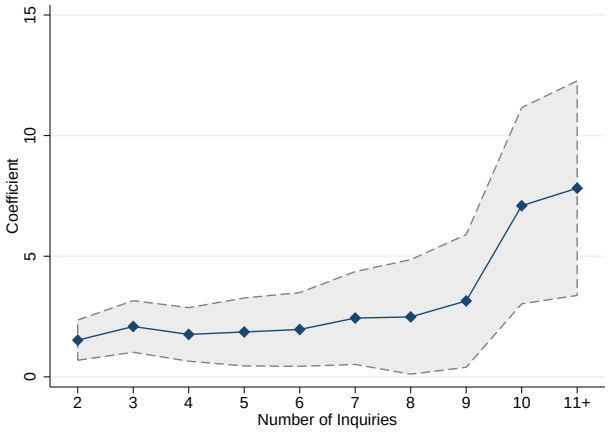
Figure 26: Coefficient Plots from Default Regression 2 - by Race Subsample



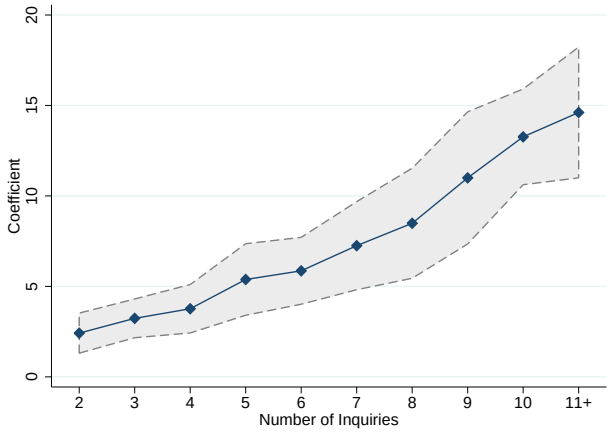
Panel A: White



Panel B: Black

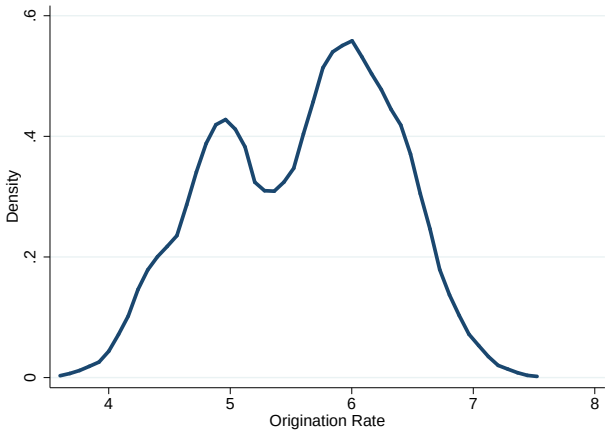


Panel C: Asian

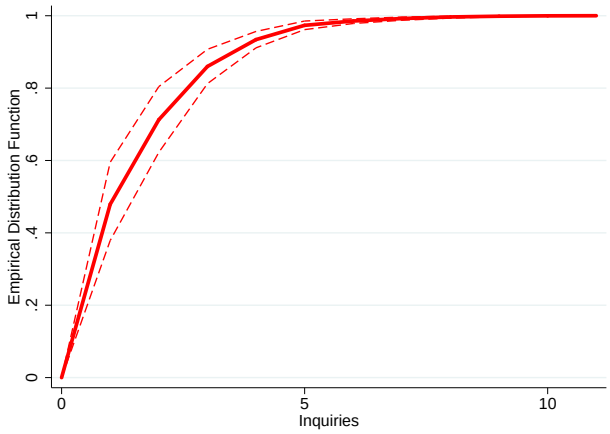


Panel D: Hispanic

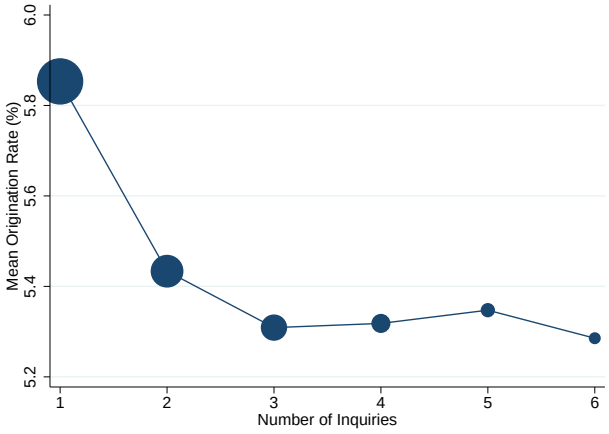
Figure 27: Search Behavior of Rarely Rejected Borrowers - Alternative Rarely Rejected Definition



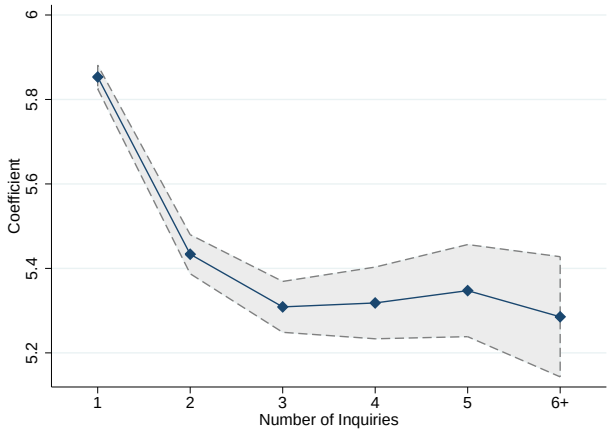
Panel A: Realized rate distribution



Panel B: Search distribution



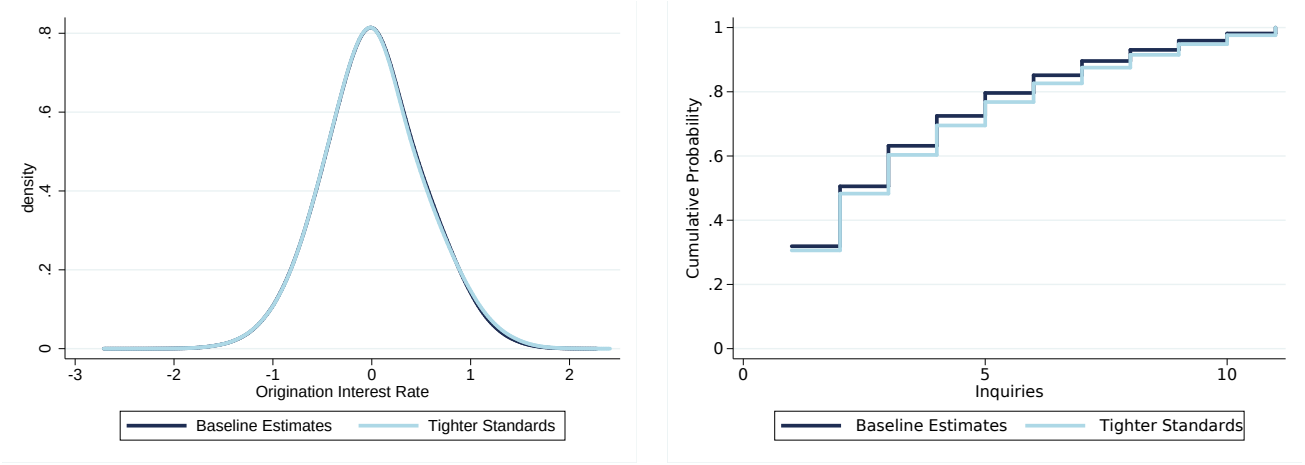
Panel C: Relationship between search and prices



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.

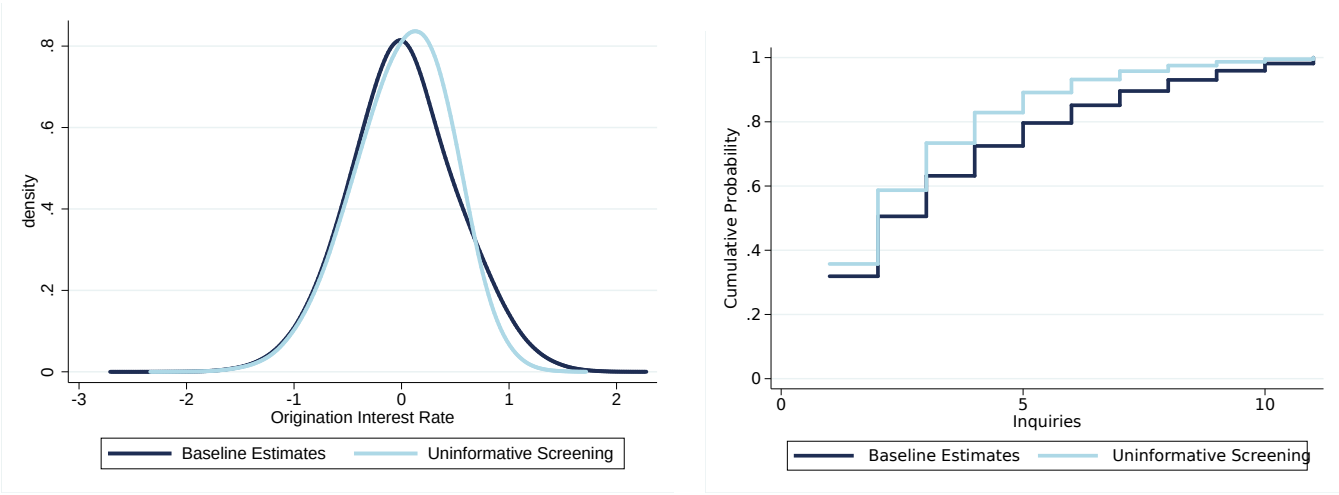
Figure 28: Realized Rate Distribution under Tighter Lending Standards Counterfactual - No Supply Adjustment



Panel A: Realized Rate Distribution Panel B: Search Distribution

Notes: Figure plots distribution of realized rates (Panel A) and of search (Panel B) 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), fixing the offered rate distribution at the baseline level.

Figure 29: Realized Rate Distribution under CRA Counterfactual - No Supply Adjustment



Panel A: Realized Rate Distribution Panel B: Search Distribution

Notes: Figure plots distribution of realized rates (Panel A) and of search (Panel B) 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, fixing the offered rate distribution at the baseline level. The constant application approval rate is given by the average approval probability under our baseline estimates: $\lambda p_h + (1-\lambda)p_l$.

12 Appendix 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 12. 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.

13 Appendix C: Likelihood Construction

13.1 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, and χ_{is} be an indicator for whether borrower i applied for the loan offered to her on her s^{th} search. We assume O_{is} has CDF given by $H(o)$.

We proceed using a maximum likelihood approach. We now construct the likelihood of observing our data. 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^*. We first consider the probability that the borrower originates a loan at a rate below r on her s^{th} inquiry. We have (suppressing the loan index i for legibility)

$$\begin{aligned} Pr\left\{R \leq r, S = s | z, r^*\right\} &= Pr\left\{\left(O_s \leq r \leq r^* \cap A_s = 1\right), \neg(O_1 \leq r^* \cap A_1 = 1) \cap \ldots \cap \neg(O_{s-1} \leq r^* \cap A_{s-1} = 1) | z, r^*\right\} \\ &= Pr\left\{\left(O_s \leq r \leq r^* \cap A_s = 1\right) | z, r^*\right\} \left(Pr\left\{\neg(O \leq r^* \cap A = 1) | z, r^*\right\}\right)^{s-1} \\ &= \mathbf{1}\left\{r \leq r^*\right\} \cdot p_z H(r) \left(1 - p_z H(r^*)\right)^{s-1} \end{aligned}$$

where $\neg$ represents logical negation. The second 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. Integrating out the condition on the borrowers' reservation rate and type yields the likelihood function for the joint distribution of origination rates and search:

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

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 follow a proportional hazard model. 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 $1-x_z^{t/T}$, regardless of prior search or acceptance. Thus we may write the likelihood of the joint distribution of our loan data $(S_i,R_i,D_i|A_i=1,\chi_i=1;t,T)$ as follows:

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

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 an application having s inquiries: $Pr\{S = s | \chi_s = 1\}$. 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

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

It is straightforward to show that the numerator may be written as

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

It remains to derive $Pr\{\chi_s = 1\}$, which is the probability that the s^{th} inquiry enters our application data. 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

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

The probability that we observe an application with s inquiries is thus the probability of observing the s^{th} inquiry, divided by the total probability of observing any inquiry up to $\tilde{S}$:

$$
Pr\{s-1 \text{ failed inquiries} | \chi_s = 1\} = \frac{Pr\{s-1 \text{ failed inquiries} \cap \chi_s = 1\}}{\sum_{1 \le s' \le \tilde{S}} Pr\{s'-1 \text{ failed inquiries} \cap \chi_{s'} = 1\}} \qquad (12)
$$

The denominator may be written as

$$\lambda \int H(r^*) \sum_{1 \leq s' \leq \tilde{S}} (1 - p_h H(r^*))^{s'-1} \, dF_h(r^*)$$
$$+ \quad (1-\lambda) \int H(r^*) \sum_{1 \leq s' \leq \tilde{S}} (1 - p_l H(r^*))^{s'-1} \, dF_l(r^*)$$

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

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

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.[36]

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 $G(c)$. Given this assumption, the recovery of search cost and offered rate distributions suffices to estimate the reservation rate for high and low types. To see this, recall that a type z borrower has the following relationship between their search cost c and reservation rate r^*

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

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

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

[36] 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.

If $\psi_z(r^*)$ is easily calculable, our estimation problem now only requires the estimation of two distributions - that of the borrower search costs and offered rates - rather than three. Here, 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

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

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 2. A detailed description of this construction is provided in Appendix 14.1.

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 $\theta \equiv \{p_h, p_l, x_h, x_l, \lambda, \beta_H, \beta_C\}$

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

where $l(R_i, D_i, S_i|\theta)$ is given by equation 10, and $l(S_i = s|\theta)$ 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. 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.

13.2 Calculating Market Shares

To construct the market share of type z individuals, $q_z(r)$, 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:

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

13.3 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 distribution of 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.[37] As a result, letting S be the size of the market,

[37] 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

the expected profits from making a loan at rate r are

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

Additionally, recall from equations 5 and 14 that the market share of type z individuals may be written as:

$$q_z(r) = \int_r^\infty \frac{f_z(r^*)}{H(r^*)}dr^* \qquad \text{and} \qquad \frac{dq_z(r)}{dr} = -\frac{f_z(r)}{H(r)} < 0$$

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.

14 Appendix D: Computational Details

14.1 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:

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

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

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

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 $\beta_G \equiv \left\{\mu_n^G, \sigma_n^G, \pi_n^G\right\}_{n=1}^N$.

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 may be calculated easily 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 $H(o)$.

14.2 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.

The challenge to such analysis is clear. Both the market share equations (5) and (14) and reservation rate distributions depend on the distribution of offered rates in the market. Therefore, 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 the establishment of equilibrium offered rate distributions necessitates the solution of 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 calculate the distance between $H(r;\beta_H)$ and $\hat{H}(r;\beta_H)$, by summing the squared distance between the two distributions' mean and variance. By minimizing this distance with respect to β_H, we may calculate close approximations to the equilibrium offered rate $H(r;\beta_H)$. 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.[38] 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

[38] 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.

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.