Searching for Approval

Sumit Agarwal, John Grigsby, Ali Hortacsu, Gregor Matvos, Amit Seru, and Vincent Yao∗

**Preliminary: Please do not cite without authors’ permission**

October, 2019

Abstract

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

∗We thank Tobias Salz, as well as seminar participants at the 2019 American Economic Association Meetings, University of Chicago, University of Pennsylvania, Columbia University, Northwestern University, UT Austin and Stanford GSB for useful comments. Kyongbae Kim provided exceptional research assistance. Authors’ contact information: Agrawal: ushakri@yahoo.com; Grigsby: jgrigsby@uchicago.edu, Hortacsu: hortacsu@uchicago.edu; Matvos: gregor.matvos@mccombs.utexas.edu; Seru: aseru@stanford.edu; Yao: wyao2@gsu.edu.
1 Introduction

Consumer credit markets exhibit substantial price dispersion. Borrowers with similar characteristics obtain credit with substantially different interest rates or fees in both mortgage markets (Gurun et al. 2016, Allen et al. 2014, Hall and Woodward 2012), and credit card markets (Ausubel 1991, Calem and Mester 1995, Stango and Zinman 2013). Moreover, the cost of credit varies widely along observable dimensions, and tends to be higher for less educated, poor, low credit score (subprime), or minority borrowers. A leading explanation of these facts is consumer search. Less sophisticated borrowers search less, and consequently settle for 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 mainly infers search behavior from the price distribution, or, in rare cases, measures search behavior from surveys, which are rarely linked to consumers’ choices.

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

In order to understand search behavior in the mortgage market, we focus on a distinction between credit markets and markets for non-financial consumer goods. While sellers of most consumer products do not have payoffs that depend on who purchases their product, credit providers’ profits depend directly on the probability that their customers repay their loans. As a result, lenders evaluate borrowers’ creditworthiness, and approve or reject customers based on this evaluation. If an application is rejected, customers must search for a mortgage with another lender. This approval process differentiates search in the mortgage market from search in markets for standard goods, such as books or autos. Indeed, this approval process is not limited to the mortgage market. It is a common feature in obtaining a credit card, student and small business loans, or in auto loans, and a similar process takes place in the insurance industry, where applicants are screened for underlying risks. However, most empirical search models of financial markets do not account for this important institutional detail.

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 accounting 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. Contrary to the predictions of the canonical search model of Carlson and McAfee (1983), the average realized interest rate does not monotonically decline with search. Borrowers, who search a lot, obtain higher rate 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 and origination date. This
result cannot be generated by canonical search models, which have been applied to the market for books (De Los Santos et al. 2012), mutual funds (Hortacsu and Syverson, 2004; Roussanov et al., 2018), and the mortgage market (Allen et al., 2014; Allen et al., 2018).

In addition, we show that borrowers who search are more likely to be delinquent or default on the loans ex post, even conditional on very detailed ex ante measures of their creditworthiness, such as FICO, LTV, DTI. Finally, using novel approval data, we show a robust negative relationship between the probability of mortgage approval and the number of searches. This is the case even in a narrow time window during which prior borrower searches are not observed by the lender. Standard search models, lacking any notion of creditworthiness or application rejection, are unable to match the fact that borrowers who search more are more likely to fall into delinquency and have their applications rejected.

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

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

First, one cannot infer consumers’ 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 generates endogenous adverse selection. Moreover, adverse selection arises even when creditworthy borrowers have a higher willingness-to-pay for loans, which would result in advantageous selection in standard frameworks. Since low creditworthiness borrowers behave as if they have higher search costs, such borrowers are endogenously more likely to take up expensive mortgages, leading firms charging higher prices to have
lower quality borrowers on average.

We estimate the model using a maximum likelihood procedure which utilizes the joint distribution of search, interest rates, default, and application approval. The estimated model successfully replicates the qualitative and quantitative patterns we observe in the data. First, frequent-searchers pay higher interest rates because they are, on average, of low unobserved 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. In other words, they pay high rates because of their search behavior, not directly due their low creditworthiness.

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

As further validation of the mechanism proposed by the paper, we examine a population of borrowers who face almost no possibility of their mortgage application being rejected. These borrowers, with approval rates of almost 98.75%, differ substantially from the overall population, whose rejection probability is approximately 18%. Our model predicts that, in the absence of any possibility of application rejection, borrowers should behave as if they were searching in any standard consumer goods market, such as the market for books. As a result, rarely-rejected borrowers sort only on search costs, leading borrowers who search more to obtain cheaper mortgages - there should be a negative relationship between search and realized prices. Strikingly, the data show that mortgage origination rates are monotonically decreasing in the frequency of search for the population of rarely-rejected borrowers, standing in stark contrast to the patterns for the population at large. These results provide additional support for our model, and suggest that the non-negative relationship between search and mortgage rates for the overall population is indeed driven by the approval process rather than some other unobservable borrower characteristic.

The model estimates imply that screening is quite informative: high types are approved with a probability which is 81 percentage points higher than low types. However, low type borrowers compose the majority of the population. Consistent with the existing literature on search in mortgage markets, the mean search cost is large, with each additional search being equivalent to paying an additional 29.7 basis points on a loan, which equates to a cost of $10,603 over the life of an average-sized loan in our sample. In addition, search costs have a standard deviation of 11.8 basis points. Estimating the model by subsamples, we find the intuitive result 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.

The estimated model permits unique counterfactuals measuring the passthrough of credit standards to pricing. 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. Borrowers internalize the tighter lending standards into their reservation
price, and thus 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, computing the equilibrium supply side response to change in borrowers’ search behavior, and hence change in demand elasticity, this increase in reservation rates induces lenders to increase their offered rates, pushing realized prices yet higher. With the supply side response, we estimate that tighter lending standards during the crisis increased average mortgage rates by 25.4bp.

We next estimate the impact that screening has on the market by considering the opposite scenario in which all borrowers had their applications accepted with certainty. In this case, our model collapses to the baseline search model. Removing screening from the model reduces mean realized rates by 25.4 basis points. What’s more, removing screening from the model reduces the standard deviation of realized rates by 29%, suggesting that screening accounts for approximately a third of the variation in GSE mortgage rates. In addition, the removal of screening roughly halves borrowers’ total outlay on search costs, but leads to an upper bound decline of bank profits of $36 billion. Screening is therefore a key feature of the mortgage market, and significantly contributes to price variation, search costs, and bank profits.

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 from the discriminated group 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 2bp 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 26.9bp.

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 screening and credit approval process is therefore critical for understanding how consumers search for credit products, and more broadly, products in which the seller’s payoff depends on buyer’s characteristics, such as insurance.

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, Allen et al. 2018). 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. Section 3 describes the data used in our empirical analysis. In section 4, we present the basic facts of search in mortgage markets, as well as the relationship between search and prices. We present our model of search with screening in section 5. Section 6 presents additional evidence in support of the screening mechanism central to our model, such as the relationship between search and both delinquency and approval probabilities. We describe the estimation of our model in section 7 and report its results. Finally, section 8 presents our counterfactual analyses. Section 9 concludes.

2 Credit Application Process and Inquiries

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

\footnotetext[1]{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].}
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. In other words, during this period, potential lenders cannot observe whether a borrower applied for any other loans. 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 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, potential borrowers searching for a mortgage are entitled for a “shopping window” of 45 days during which multiple credit checks from mortgage lenders are recorded on the searchers’ credit report as a single inquiry.² This shopping window starts with the first credit check by a mortgage lender, and applies only to credit checks from mortgage lenders and brokers; credit card related and other inquiries are registered separately.

Third, as mentioned, similar formal credit 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 part of the “total inquiries” in the credit bureau data. Several observations suggest that these non-mortgage inquiries will not pollute the interpretation of total inquiries as mortgage search. First, 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 inquiries³ 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 preceding 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

³As determined by the credit bureau.
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 applications for mortgages intended to purchase or
refinance a single family property, from 2001 to 2013. The loans are originated by a variety of lenders and conform to
GSE standards. We consider only loan applications with a single applicant, because they tend to have cleaner search
histories at the time of application. The sample contains common underwriting variables, including borrower credit
score, backend debt-to-income (DTI) ratio, loan-to-value (LTV) ratio of the mortgage, mortgage contract choice,
loan purpose (purchase vs refinancing), occupancy (primary residence vs investment property), application date and
property location, for both approved and rejected loan applications.

Our second dataset contains approximately 1.3 million mortgages that were originated between 2001 and 2011. At
origination, we observe the borrower’s credit score, the LTV ratio, the loan characteristics (origination balance, note
rate, and term), the backend DTI ratio, whether the loan was originated through a broker, loan purpose, occupancy,
and the location of the mortgaged property (zip code, MSA and state). In addition, we also have information on some
of borrower’s demographics, including years of school, age, gender and their monthly income at origination. 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 heterogeneity in observed creditworthiness in our pool.
The standard deviation of FICO scores is 62.5 in the loan-level dataset, and 71.6 in the application dataset. We see
similarly large standard deviations in both CLTV and DTI ratios. Indeed, these loans are not without credit risk:
15.95% had entered default by January, 2015.

---

4To limit the influence of outliers, we winsorize applications and loans lying above the 99th percentile of inquiries, interest rates, DTI,
or LTV ratios.
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 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. This has been shown in the US subprime market (Gurun et al. 2016), as well as in Canada (Allen et al. 2014). Borrowers pay substantially different mortgage rates in our sample as well, even after adjusting for points and fees. We present the full distribution of rates across three origination time periods in Figure 1A, showing substantial rate dispersion. Figure 1B presents interest rates for three different FICO based creditworthiness subsets. There is substantial mortgage rate dispersion within every subset, with interest rates differing over 3 percentage points (pp) within each group. These differences are costly. The average loan in our data is originated for $169 thousand, so each pp represents an additional $1,200 in interest expense every year for a 30-year fixed rate mortgage (FRM).

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

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

in which \( r_{itm} \) represents the origination rate of borrower \( i \) at time \( t \) in market \( m \). \( X_i \) are the borrower and loan characteristics, such as FICO score, LTV, DTI, income, years of education, the type of the mortgage, and whether the borrower is an investor. It is worth reiterating that we observe the actual characteristics, rather than a noisy proxy derived from borrowers’ locations, as is used by the majority of mortgage research. In order to compare borrowers

---

in the same market at the same point in time, we condition on state fixed effects $\mu_m$, and on time fixed effects $\mu_t$. Our data set was 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 1C) is compressed relative to the distribution of raw origination rates, suggesting that at least some of the dispersion in rates is driven by observed borrower differences. However, a substantial amount of residual rate dispersion remains. A borrower at the 10th percentile of the distribution pays an origination rate that is 0.9pp lower than that paid by the borrower at the 90th percentile of the distribution. At the average loan amount of $169$ thousand, this difference results in $1,080$ larger mortgage cost per year. Our estimates of residual price dispersion of 41bp are similar to 50bp found in Allen et al. (2014). Meanwhile Gurun et al. (2016) find a coefficient of variation of 0.23 and 0.19 in their data on fixed- and adjustable-rate mortgages, respectively, compared with 0.15 in our data. Overall, borrowers with the same characteristics, in the same market, borrowing from the same lender at the same point in time pay substantially different mortgage rates.

4.2 Borrower Search, Sophistication, and Creditworthiness

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

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

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 2C, and across educational attainment in Figure 2D. Consistent with intuition, the most educated borrowers search most, but the difference is slight and statistically insignificant. FICO, which measures creditworthiness, is among the strongest predictors of search: low FICO scores (below 620) search substantially more than borrowers with high FICO scores (above 720). These simple facts suggest that differences in creditworthiness play an important role in understanding

---

7To test whether brand preferences or non-price aspects of a particular lender account for the observed price dispersion, we add lender×origination quarter fixed effects. Adding these increases the $R^2$ from 0.80 to 0.81 and reduces the residual standard deviation from 41bp to 40bp.

8The 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.
search in the mortgage market.

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

\[ s_{itm} = \beta X_i + \mu_{mt} + \varepsilon_{itm} \]  

(1)

in which \( i \) indexes the mortgage applicant or borrower in market \( m \) at time \( t \). The dependent variable \( s_{itm} \) is the number of inquiries, or an indicator that the borrower belongs to the \( n^{th} \) quartile of search, scaled by 100 for legibility. We examine the conditional correlation between search and borrower characteristics, such as their FICO score, education, income and race. To ensure that the correlation between characteristics and search is not driven by local or aggregate conditions, we include the location-time fixed effect \( \mu_{mt} \). Any differences in the regulatory environment are also absorbed by the location fixed effect. We present the results in Table 3. Panel A reports estimates for our sample of mortgage applicants, while Panel B reports estimates for our sample of mortgage borrowers. Borrower characteristics such as education and race are correlated with the amount of search, but the simple correlations are not consistent with the intuition that sophisticated borrowers search more. More critical to the argument, more creditworthy borrowers search less, even conditional on other characteristics, suggesting an important role for creditworthiness in understanding consumer search behavior. Indeed, a borrower with a FICO score which is one standard deviation above the mean has 3.8 fewer inquiries on average in the application data, and 0.39 fewer inquiries in the realized loan data, conditional on other observable characteristics. However, college educated borrowers, traditionally considered sophisticated, have 0.11 fewer inquiries than non-college borrowers at the time of mortgage origination.

4.3 Do Borrowers who search more obtain cheaper mortgages?

The benchmark consumer search model suggests that search and transacted prices are negatively correlated, 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 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 more expensive products on average.

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

Figure 3 cuts the data on several other dimensions, which may drive search and mortgage pricing - FICO, race, income, and education - and plot the relationship between search and interest rates for each group. The same pattern
persists for low, middle and high FICO scores, low, middle and highly educated populations, for black, white, and Hispanic borrowers, as well as for low, middle, and high income borrowers. 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 account for conditional correlations, we next explore the relationship between mortgage rates and search in a regression framework, in which we can control for differences across markets, as well as borrower and mortgage characteristics. We estimate the following regression

\[ r_{itm} = \sum_{s \geq 2} \beta_s 1\{s_i = s\} + \mu_t + \mu_m + \beta X_i + \epsilon_{itm} \]

(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 \( \beta_s \) 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), investor status, product type (ARM vs FRM, purchase vs refinance), and backend DTI ratio. To ensure that our results are not driven by local supply or demand conditions, we include the time fixed effect \( \mu_t \) and location fixed effect \( \mu_m \). These fixed effects will also absorb any aggregate fluctuations, such as changes in the risk premia, or persistent differences across markets. We cluster standard errors at the state x origination quarter level.

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

Finally, Appendix 13 tests robustness of these patterns by estimating equation 2 across a number of subsamples. In particular, Appendix Figure 19 estimates equation 2 controlling for a richer set of covariates, namely the set of loan-level price adjustment (LLPA) factors used by Fannie Mae. In each case, we observe a U-shaped or positive
relationship between search and interest rates in the data. Overall, the predictions from the standard search models, that more search is correlated with lower mortgage rates is rejected. We therefore develop a theory, which is able to generate these patterns.

5 Model

We now 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. While the application / rejection process can be endogenized, the goal paper is to evaluate the impact of rejections on search. Second, the model yields new testable predictions, 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 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, applications are subject to approval by the lender. Upon receiving a mortgage application, lenders can perform an in-depth credit check to obtain imperfect, but informative information on the borrower’s creditworthiness. The credit check is valuable, because creditworthiness is the 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 probability of repaying a loan in full $x_z \in \{x_h, x_l\}$, with $Pr(x_z = x_h) = \lambda$. Borrowers with high repayment ability (creditworthiness), are more likely to repay a loan than borrowers with low repayment ability: $x_h > x_l$. Creditworthiness and search costs are i.i.d across consumers and types. A consumer $iz$’s utility from obtaining a mortgage from lender $j$ at rate $r_j$ is:

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

Consumers prefer loans with lower interest rates. Further, to illustrate that standard adverse/advantageous selection does not drive our results, we allow consumers with different creditworthiness to have different preferences over

---

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

10The i.i.d. assumption is useful to cleanly separate the effect of search costs from creditworthiness.
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.\footnote{Indeed, there is no large difference in the relationship between search and interest rates for borrowers who default on their loans ex post compared with those who do not: re-estimating equation Equation 1 for these two populations yields similar estimates of $\beta$. See Appendix Figure 15.}

5.1.2 Lenders and Mortgage Approval

Lenders post mortgage interest rates. Lenders choose from a menu of $K$ discrete potential rates to offer, $r_k \in \{r_1, \ldots, r_K\}$.\footnote{We transform the problem of choosing an offered rate 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.} 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, as well as regulatory and administrative costs.

We depart from the standard sequential search model by assuming that the potential borrower observes her creditworthiness, $x_z$, but the lender does not. Before obtaining a mortgage, the borrower is subject to an application approval process. The lender carries out an in-depth check of applicants’ creditworthiness, which generates an informative, but imperfect, signal $s_i \in \{s_h, s_l\}$. If the borrower is of repayment ability $x_z$, the probability that she is revealed as a high type is $p = Pr (s_h | x_z)$. The in-depth review is informative $p_h \geq p_l$, so high repayment ability borrowers are more likely to be revealed as such. We assume that applications generating signal $s_h$ (indicating the borrower is high type) are approved, while those generating $s_l$ are rejected. We nest the benchmark model without approvals by assuming screening is uninformative, $p_h = p_l = p$. Consistent with the institutional environment, lenders do not observe borrowers’ prior search.\footnote{Observing prior search would not eliminate adverse selection induced by rejections, since the lenders do not know whether a borrower’s search continues because of rejections or in search of better rates.}

5.2 Consumer search (Demand)

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

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:

$$c_i \leq \int_{L}^{r} \frac{Pr(s_h|x_z)((-\bar{r} + \sigma x_z) - (-r + \sigma x_z))dH(\bar{r})}{pr.\,\text{approval}}$$

$$c_i \leq p_z \int_{L}^{r} (r - \bar{r})dH(\bar{r})$$

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

Denote by $r^*_{iz}$ the highest rate that the borrower with search cost $c_i$ and repayment type $z$ would accept. At this rate, the borrower is indifferent between searching further and accepting the mortgage:

$$c_i = p_z \int_{L}^{r^*_{iz}} (r^*_{iz} - \bar{r})dH(\bar{r})$$

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

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

$$\frac{c_i}{p_z} = \int_{L}^{r^*_{iz}} (r^*_{iz} - r)dH(r)$$

(4)

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

### 5.2.1 Approval Process Induced Adverse Selection

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

\[
p_h \int_{r_h^*}^{\infty} (r_h^* - r) \, dH(r) = p_l \int_{r_l^*}^{\infty} (r_l^* - r) \, dH(r).
\]

\(p_h > p_l\) implies that \(r_h^* < r_l^*\). That is, less creditworthy borrowers are willing to accept higher mortgage rates than more creditworthy borrowers with the same search cost. For adverse selection to occur, the approval process must be informative. Despite underlying asymmetric information, if rejection rates are the same for both types of borrowers, \(p_l = p_h\), we revert to a model with no adverse selection.\(^{16}\)

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

### 5.3 Interest rate setting (Supply)

Lenders only accept borrowers who apply for their loan and whose credit check generates a positive signal \(s_h\). Because borrowers sort, setting the interest rate affects both the probability of repayment on their pool of mortgages, and the expected quantity of mortgages the lender will underwrite, \(S (\lambda q_h (r_j) + (1 - \lambda) q_l (r_j))\), where \(S\) is the total size of the market and \(q_z(r)\) is the share of the type \(z\) market a bank charging a rate \(r\) can expect. We assume that screening is valuable, which is consistent with observing rejected applications in the mortgage market. The expected profits from charging an interest rate \(r\) are thus:\(^{17}\)

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

\(^{16}\)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 \(\sigma\) operates on the extensive margin of obtaining a mortgage in the first place, and therefore drops out of the search problem.

\(^{17}\)The profit function is specified in terms of percentage points of interest. In our empirical application, we residualize observed interest rates against borrower characteristics, so that the interest rate \(r\) may take on positive or negative values. One may thus interpret \(\Pi(r)\) as the excess return, in percentage points, that a lender may earn if it charges a rate \(r\) percentage points above the average realized rate for an observably equivalent borrower in the market.
Letting $f_z(r^*)$ be the density of reservation interest rates for borrowers of type $z$, Appendix 14.2 shows that the market share of type $z$ individuals that a bank offering rate $r$ captures may be expressed as

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

(5)

Intuitively, undirected search implies that a lender charging a rate $r$ obtains a fraction $1/H(r^*)$ of the market share for borrowers with reservation rate $r^*$, and can capture the mass of individuals with reservation rates above $r$. Each lender faces an additional idiosyncratic profit shock to charging specific rates $\xi_{j,k}$, which are i.i.d. and distributed according to a Type 1 Extreme Value (T1EV) distribution with scale factor $\sigma_\xi$. 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 $\sigma_\xi$, the probability that rate $r_k$ maximizes the lender’s profit is:

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

(6)

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

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

$$\frac{d\mathbb{E}[\Pi(r|m)]}{dr} = q(r) \left( \mathbb{E}\left[\hat{x}_k|\hat{r}, s_h\right] \right) + \frac{\partial q}{\partial r}(r) \left( r\mathbb{E}\left[\hat{x}_k|\hat{r}, s_h\right] - m \right) + q(r) r \frac{\partial \mathbb{E}\left[\hat{x}_k|\hat{r}, s_h\right]}{\partial r}$$

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 $\frac{\partial q}{\partial r} \leq 0$, because the marginal borrowers now choose to keep searching instead of accepting the mortgage. This downward-sloping residual demand curve highlights banks’ market power in this setting. The profits lost on each borrower are $(r\mathbb{E}\left[\hat{x}_k|\hat{r}, s_h\right] - m) \geq 0$. The second cost of increasing mortgage rates is that a higher interest rate attracts a weakly worse pool of borrowers,

\[18\] These assumptions come into play when computing counterfactuals, and do not play a role in the qualitative predictions of the model.
\[ \frac{\partial E[\tilde{x}_k|r,s_h]}{\partial r} \leq 0. \] The borrower pool for firms with high rates is worse because more creditworthy borrowers have lower reservation rates, and are therefore less likely to accept a mortgage when the price increases. This last component changes lenders’ pricing incentives relative to a standard search model. In the benchmark model the search behavior and reservation rates are independent of borrowers’ creditworthiness, which implies that \[ \frac{\partial E[\tilde{x}_k|r,s_h]}{\partial r} = 0. \] Therefore, approvals change the lenders’ pricing incentives on the margin by introducing adverse selection, which decreases incentives to raise mortgage rates on the margin.

5.4 Equilibrium

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

\[
\int q_z(r) dH(r) = 1 \quad z \in l, h
\]

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 15.2.

5.5 Model predictions

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

5.5.1 Benchmark: All mortgages are approved

As the probability of approval for both types goes to one, the model reverts to a standard search model without the approval process.\(^{19}\) We show that this benchmark model’s predictions are inconsistent with the relationship between search and rates documented 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:

\[
c_i = \int_{\xi}^{x_i} (r^*_z - r) dH(r)
\]

\(^{19}\)In fact, it is sufficient that \( p_l = p_h = p. \)
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 $\lambda$, 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: $Pr(\text{Search again}) = 1 - H(r^*_iz)$. Then the probability that a borrower with a reservation rate $r^*$ searches more than $s$ times is:

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

Low search cost (financially savvy) customers, have lower reservation rates, $r^*$, and are therefore more likely to search. Because they have lower reservation rates, their average interest rate on accepted mortgages is lower. This induces a negative relationship between search and average interest rates, as illustrated in Appendix Figure 16 for a simulated sample of borrowers. Overall, the relationship between average rates and search in the data rejects this prediction.

### 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. Recall that lenders do not observe borrowers’ search, and therefore cannot condition mortgage pricing on search. The relationship between search and pricing therefore arises because of borrower search behavior. The possibility of application rejection creates two reasons for a borrowers to continue to search. First, there exists the standard reason for continued search: a borrower might draw a mortgage with an interest rate above their reservation rate, $r > r^*_iz$, and so choose not to apply for the mortgage. Alternatively, the borrower might discover a mortgage with $r \leq r^*_iz$ for which they apply, only to have her application declined. The total probability that a borrower searches again is thus:

$$Pr(\text{Search again}) = 1 - Pr(r < r^*_iz) + Pr(r < r^*_iz)(1 - p_z)$$

$$= 1 - H(r^*_iz)p_z.$$

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

$$Pr(S_{iz} > s|r^*_iz = r^*) = (1 - p_zH(r^*))^s$$

19
The two forces work in opposite directions. Less creditworthy borrowers are more willing to accept higher rates – \( H(r_t^*) \) is higher – which pushes them to search less. However, less creditworthy borrowers are also more likely have their application rejected if they find a mortgage with a low enough rate, urging more search. If the latter force is strong enough, high type borrowers disappear from the population of searchers faster than low type borrowers. To illustrate this, we simulate a search process with highly informative screening, and plot the results in Figure 5. Panel C presents the share of high types left in the population at each level of search. With a strong screening technology, only low type individuals remain searching at the highest levels of search, while high type individuals drop out of the sample as they find acceptable mortgages.

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

5.5.3 Default, Approval, and Search

Our model predicts a direct relationship between the equilibrium amount of search, mortgage approvals, and mortgage default, which we later test in the data. These predictions differ from the baseline model, in which approvals are not informative, the default and approval probabilities are independent of the number of inquiries. Our model predicts that an informed approval process results in a decline in the average quality of the borrower pool, as the number of searchers increases (Figure 5C). Formally, defining \( \tilde{\lambda}(s) \) to be the share of high type borrowers among loans realized after \( s \) inquiries. The average default rate of borrowers with \( s \) inquiries should be \( \tilde{\lambda}(s)(1 - x_h) + \left( 1 - \tilde{\lambda}(s) \right) (1 - x_l) \).

Since \( \tilde{\lambda}(s) \) is declining in \( s \) and \( x_h > x_l \), borrowers with a large number of inquiries should be less likely to repay the lender ex post. Figure 5E illustrates the relationship between inquiries and repayment behavior for our simulated set of borrowers in our scenario with highly informative screening.

Similarly, the probability that a loan application is accepted for a borrower with \( s \) searches as \( \tilde{\lambda}(s)p_h + \left( 1 - \tilde{\lambda}(s) \right) p_l \).

Since the type of a borrower who applies for a mortgage after many searches is of lower average quality, those with high inquiry counts are more likely to be rejected upon the in-depth review. As a result, lenders are more likely to reject borrowers who search more, even if they cannot observe the number of searches. Figure 5F shows this decreasing relationship between application approval probability and inquiry counts for our simulated data. These predictions further distinguish our model from the baseline search model without an informative approval process.

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 informative screening from a benchmark model without approvals. As we show in Section 4.3, the relationship between search and prices (prediction 3) is consistent with our model. We now test our model by verifying that predictions 4 through 6 are also observed.

6 Additional Empirical Evidence

6.1 Loan Performance and Search

Our model predicts that less creditworthy borrowers search more in equilibrium, leading to positive relationship between search and ex-post default rates. Figure 6 plots the annualized default rate against the number of inquiries on record for all borrowers in our sample. 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 ex-post less creditworthy.

High-inquiry borrowers may simply be of lower credit quality on dimensions observable to the lender. Indeed, Figure 2C and Table 3 show that low FICO borrowers do indeed search more. To test whether frequent searchers are more likely to default even conditional on observables, we estimate the following linear regression:

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

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, scaled by 100 for legibility. The independent variable of interest is the amount of search the borrower undertook before taking up a mortgage, $s_i$. The coefficients of interest $\beta_s$ measure the difference in default probability for borrowers who search $s$ times compared with those who search just once. As with our interest rate regressions, we control for observable characteristics, and include a time fixed effect $\mu_t$ and location fixed effect $\mu_m$. As before, these

20Our 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 $\tau$ years before we observe loan performance, the annualized default rate $\tilde{d}$ would solve $1 - \tilde{y} = (1 - d)^{\tau}$. 

21
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, $\beta_s$, in Panel C of Figure 6. Consistent with our predictions, borrowers who search more are more likely to default or become delinquent on their loans, even conditional on observable characteristics. A borrower who searches 5 times is approximately 5 percentage points more likely to have defaulted on their mortgage as of January 2015 than is a borrower with 1 inquiry, conditional on observables. This positive relationship between search and default probabilities is highly robust. We re-estimate the specification in sub-populations of low, middle and high FICO borrowers, low, middle and highly educated populations, for black, white, and Hispanic borrowers, as well as for low, middle, and high income borrowers (Panels D-F of Figure 6). Across all sub-samples, the data supports our model’s prediction that more frequent searchers are on average less creditworthy than infrequent searchers, even conditional on observable characteristics.

6.2 Search and Approvals

Central to our model’s predictions is the borrower approval process. The model predicts that the borrower pool of frequent searchers contains more low creditworthy types, who are more likely to be rejected following an in-depth credit check. Using our application-level dataset, we are uniquely able to test this implication of our model. Since search is measured within the credit bureau’s “shopping window,” the borrower’s search history is not observed by the lenders. In other words, the lenders cannot condition approvals on prior search.

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

$$ a_{itm} = \alpha + \sum_{s \geq 2} \beta_s 1\{s = s\} + \mu_t + \mu_m + \beta X_i + \varepsilon_{itm} $$ (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 100, if the application was accepted, and 0 otherwise. Again, the coefficients of interest $\beta_s$ 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, as well as location and time fixed effects. The coefficients of interest are presented in Figure 8. Even controlling for observable loan and borrower characteristics, borrowers who search more are less likely to have their application accepted. This pattern holds across our three borrower FICO score buckets, as shown in Figure 8. The data therefore support the model’s prediction that borrowers who search more more are less likely approved for mortgages, conditional on observables.

In summary, borrowers who search more are of lower average quality in two separate datasets and along two dimensions – default and application acceptance probability. The benchmark search model, in which borrowers differ only in their search cost, would predict no relationship between search and average borrower creditworthiness. 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.

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

We select borrowers whose mortgage applications are rejected very rarely. We use all borrower, mortgage, location, and time characteristics to predict the probability that an application is accepted by estimating a logistic regression. Borrowers are said to be rarely-rejected if their predicted approval probability is greater than 97.5%. The average 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.\footnote{In Appendix Figure 20 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%.}

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

7 Model Estimation and Counterfactual Analysis

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

7.1 Estimation

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

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

We estimate the model using maximum likelihood using our two datasets. The first dataset contains information on mortgage applications and the distribution of inquiry counts conditional on application. The second dataset is at the loan-level, and reports the origination interest rate, loan performance, and inquiry count at the time of application. That is, we observe the joint distribution of search, rates, and default, $(S_i, R_i, D_i)$, as well as a number of observable loan and borrower characteristics. To ensure comparability of realized loans in our estimation, we residualize observed rates against observable characteristics following regression equation 2. The identification problem may be stated as follows: given the distribution of $S_i$ conditional on application, and the joint distribution of $(S_i, R_i, D_i)$ conditional
on application approval, we must uniquely recover the set of model primitives. On the consumer side, we have to recover the search cost distribution \( G(c) \), the share of creditworthy types in the population, \( \lambda \), and the two repayment ability parameters, \( \{x_h, x_l\} \). On the lender side, we're interested in the screening technology, \( \{p_h, p_l\} \), the costs of making loans \( m \), and the variance of the TIEV profit shocks, \( \sigma_\xi \).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 \( H(\tilde{r}) \) must align with the firms’ choice probabilities. This suggests a robust approach to estimating the supply side parameters by minimizing the distance between our maximum likelihood estimates of \( H(\tilde{r}) \) and the choice probabilities as given by equation 6. We describe the construction of likelihood functions and details of our estimation in Appendix 14, and computational details of the estimation in Appendix 15.1. In Appendix 12, we test the binary type assumption, and find little evidence for the presence of a third type.

7.2 Results

Data Fit: Despite its simplicity, the estimated model matches observed price dispersion and distribution of searches (Figure 10, Panels A and B). The model replicates an increasing relationship between interest rates and search, and search and default documented in sections 4 and 6 (Figure 10, Panels C and D).23

Screening Technology and Adverse Selection: Our maximum likelihood estimates are reported in Table 5. Our estimates suggest that most potential borrowers, 73%, are of low type: 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 that lenders make few mistakes when screening high types: \( p_h \) is close to 1, so these borrowers rarely generate a bad credit signal. That is intuitive, since a bad credit check generally requires the revelation of bad information. However, the screening process is imperfect: \( p_l \) of 19% suggests that in 19% of cases lenders’ do not uncover the bad information on low types. On the other hand, these estimates imply that the rejection probability for bad borrowers is over 80%.

The difference between \( p_h \) and \( p_l \) of 0.807 suggest that the screening technology is very informative, and applying it results in large cost savings for lenders. A simple back of the envelope suggests that the expected loss on a bad borrower applying is lowered by approximately 81% from 34pp (one minus the expected repayment of a low type) to 19% * 34pp = 6.5pp. Therefore, given the powerful screening technology and the large benefit from successful screening, lenders find it worthwhile to screen so long as its cost is not prohibitive. As we show in Section 8.2, we estimate that the presence of this screening technology increases aggregate lender profits by $1.2 billion per year.

The informative screening technology has a large impact on the search behavior of low type borrowers, leading
to substantial 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 10E. Adverse selection is not uniform across the rate distribution: it 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 This result differs from the standard intuition that adverse selection arises at the high-end of the rate distribution because high rates select on borrowers who have a higher propensity to engage in bad behavior. In our model, very few high types choose high rate mortgages. The forces shaping the right tail of the price distribution are similar to standard search models: the sorting across rates is determined by search cost. Therefore increasing rates does not markedly change the credit quality of the borrower pool. Therefore changing rates may not change the overall amount of adverse selection in the market.

Search Costs:

The mean of the search cost distribution is estimated at 29.7bp. 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 11.8bp is smaller than 23bp in Allen et al. (2014). Furthermore, this search cost is near those estimated in the mutual fund literature, ranging from 11bp-21bp in Hortacsu and Syverson (2004) to the 39bp search cost for finding an active mutual fund in Roussanov et al. (2017).

One can translate these search costs into dollar terms using a standard mortgage calculator. Specifically, suppose a loan has origination principal \( Y \), a term of \( T \) months, and a monthly interest rate of \( r \) (i.e. one-twelfth of the annual interest rate). The standard monthly payment for this loan is given by \( y = Y \frac{(r(1 + r)^T)}{((1 + r)^T - 1)}. \)

This implies that the monthly payment on a 30-year fixed rate mortgage with principal of $170,000 and interest rate of 4% per year – the mean mortgage in the data – is $811.61. How much extra would a borrower be willing to pay in order to avoid searching one more time? If the borrower searches one additional time, she would pay the equivalent of \( c \) additional basis points of interest. Now her effective interest rate on the loan is 400 + \( c \) basis points. At the mean search cost of 29.7bp, this estimate would translate into a monthly payment increase of $29.45, or an upper bound cost of $10,603 over the term of the loan. 26

---

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

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

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

26 This estimate is an upper bound in that it assumes the mortgage is never refinanced or prepaid. In addition, we do not impose any discounting in the calculation of this upper bound cost.
The 19% acceptance rate for low creditworthiness borrowers implies they behave as if their search costs are \( \frac{1}{19\%} = 5.3 \) times higher than those of good borrowers. In other words, if one views low creditworthy borrowers through the lens of a standard search model they seem very unsophisticated. They are willing to take substantially worse mortgages than are available in the market. Our estimates suggest that accounting for rejection rates is central to understanding the search behavior of such borrowers.

**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 a borrower with a given set of characteristics. In other words, the average markup is estimated to be 1.59%. The estimate is of the same order of magnitude as 1.09% for the insured Canadian mortgage market by Allen et al. (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.

**Estimates by subsample:** Finally, we estimate our model on a variety of subsamples. The results are presented in Appendix Figure 21. Estimating on observable subsamples suffers from power issues, as some subsamples do not constitute a large share of GSE loans. Nevertheless, we believe the estimates presented here to be a useful sanity check for our results. Borrower quality tends to be increasing in FICO score and decreasing in LTV ratio, whether quality is measured according to the share of high type borrowers \( \lambda \) (Panel A), or the repayment rates of low type borrowers \( x_l \) (Panel C). The repayment rate for high type borrowers is always 1. Search costs tend to be higher and more variables for low FICO borrowers and high LTV borrowers (Panels E and F), while screening power \( p_h - p_l \) is higher for these low observable quality borrowers. Interestingly, the informativeness of the screening technology is positively correlated with the cost of misclassification \( x_h - x_l \), suggesting that banks perhaps invest more in screening borrowers with low observable quality. Meanwhile, the crisis was characterized by lower repayment probabilities, lower search costs, but also less informative screening. Although these subsample estimations should be considered merely suggestive given the issues of power, we find the intuitive nature of these results to be reassuring.

8 **Counterfactual Analyses**

We now pursue various counterfactual exercises. We first study the changes that screening

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.

We first consider the impact of tightened lending standards of the sort seen during the financial crisis. Next, we estimate the impact that informative screening has on the credit market by considering a case where all mortgage applicants are accepted. We then 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 discriminated group to pay higher interest rates on average. The results of our counterfactual exercises are summarized in Table 6. 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 15.2.

8.1 Approval Process and Search

Our first set of counterfactuals studies the impact of banks’ ability to screen and reject mortgage applications on equilibrium search, mortgage rates offered by lenders, as well as mortgages that consumers eventually obtain. We conduct two simulations. In the first, banks’ do not screen nor reject applicants. In the second simulation, banks reject applicants, but do so in an uninformed fashion. These counterfactuals have two purposes. First, they clarify the effect of introducing informed approvals into search models. Second, they allow us to investigate the effects that new screening technologies have on borrower search behavior and mortgage outcomes. The last decade has seen a tremendous rise in technology based lending (fintech), which has altered the extent of screening by lenders (Buchak et al, 2018). These counterfactuals also allow us to analyze policies, which might limit banks’ ability to screen borrowers, such as the Community Reinvestment Act (CRA) of 1977.

8.1.1 The Importance of Screening Technology: All Mortgages are Approved

We first simulate the model supposing that all mortgage applications are accepted \( p_h = p_l = 1 \). In other words, we simulate the benchmark models of search, and present the results in Table 6. Note that banks still account for observable differences between borrowers, such as FICO and LTV; they just cannot obtain a signal on borrowers type in excess of these observables. This does not remove the underlying asymmetric information, just the banks’ ability to act on it.

We show that the ability to screen borrowers, and reject those with bad signals benefit banks at the expense of consumers. Because of the search environment, even high type borrowers who are never rejected benefit from eliminating screening. Broadly, screening has two first order effects in our model. First, it forces some low type borrowers to search more after having an application rejected; importantly, it also increases their willingness to accept expensive mortgages. In response, lenders shift the distribution of reservation rates higher. This supply response induces even some high type borrowers to accept higher interest rates, even though screening does not affect them directly. The responses then further interact in equilibrium.

Removing rejection benefits consumers by reducing the mean realized rates by 2.1pp. Low types clearly benefit from this change, since they are pooled with the high types and do not incur the cost of rejection. A perhaps surprising result is that benefits, albeit smaller, accrue to high types as well. These benefits to high types arise because the mean offered rate declines by almost 2pp. As we explain below, changes in borrowers’ search behavior are the impetus behind this supply response. Interestingly, the standard deviation of offered rates increases substantially, to 1.259 from 0.547. Screening and rejections therefore increase average interest rates, but substantially decrease price
dispersion in the market.

Overall, searching because of rejections declines, while searching for better rates increases. On average, inquiries decline from 3.4 to 2.9. Applying this total search cost to the average loan in the market yields an average cost equivalent decreases from $102 per month to a total cost of $86 per month. The decline in search is driven by bad types: with the rejection probability of 81%, bad types have to search a lot just to be approved for a mortgage. Removing rejections decreases search costs induced by rejection. High types, in fact increase their search in equilibrium. Moreover, the willingness of bad types to take an expensive mortgage decline, since their de facto search cost decline.

We can better understand the equilibrium response by analyzing consumer behavior holding bank behavior fixed. Even with fixed offer rates, the transacted rates decline on average by approximately 25bp. This difference is driven entirely by the behavior of low types, who are not less willing to take high rate mortgages.\(^{27}\) This large decline arises in equilibrium through offer rate adjustments.

There are two forces leading to the decline in offer rates. First, because bad types are less willing to take high rate mortgages, banks offered rates shift to the left. Second, despite asymmetric information, adverse selection disappears, changing the pricing incentives of banks. Low and high type individuals face the same acceptance rates, so 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.\(^{28}\) 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. Both of these forces lead to a decline in offer rates, which further decreases borrowers incentives to take high rate mortgages, putting further pressure on offered rates. Since the elasticity of banks’ residual demand is large the resulting equilibrium response is substantial.\(^{29}\)

The above numbers suggest that application rejection reduces consumer surplus substantially. This does not imply, however, that welfare would be maximized were mortgage screening outlawed. After accounting for the re-optimization of offered rates, removing rejections leads to a reduction of bank profits of 1.892 percentage points on every loan. This is an enormous loss: given that $479 billion of mortgages were originated in 2017Q3,\(^{29}\) the 1.892pp reduction in profits implies that the ability to rejection applications is worth approximately $36.25 billion (4 * 0.01892 * $479 billion) per year to banks. The large size of this effect, however, should be understood with the caveat that our model does not account for the entry and exit decisions of banks. Were there a fixed cost of operation, one might expect this large decline in bank profits to be met with a fewer number of banks operating, which in turn would impact both consumer credit access and the profits of operating banks.

\(^{27}\)Our estimates suggest that high types are approved for mortgages almost certainly, \(p_h = 1\). With a fixed offer rate, their behavior does not change.

\(^{28}\)The intuition from a standard search model dominates, so that the relationship between average prices and search becomes downward sloping, and there are flat relationships between rates, default, and search (Figure 13C).

8.2 Uninformative Approvals:

We next introduce a more realistic benchmark to simulate improvements in bank’s ability to screen borrowers. We recompute the equilibrium by again removing lenders’ ability to screen: they cannot obtain a signal on borrowers type in excess of these observables. However, we assume the probability of mortgage acceptance for both high and low type individuals is equalized.\footnote{In order to maintain the same overall application acceptance probability as is observed in our data, we set } As before, the underlying asymmetric information remains in place, we just remove the banks’ ability to act on it.

We think this counterfactual is informative about the consequences of introducing technology, which allows lenders to better screen borrowers. The mortgage market has seen an explosion of algorithmic lending, with the market share of fintech lenders reaching 12\text{pp} by 2017 (Buchak et al 2018). Second, the counterfactual allows us to investigate the consequences of The Community Reinvestment Act (CRA) of 1977. The CRA requires banks to improve credit access of low socioeconomic status neighborhoods. At the same time banks are required to lend in a safe and sound manner, potentially conflicting with the first goal. An extensive literature has investigated the extent to which the CRA has increased both credit access and the riskiness of lending (Agrawal et al. 2012, Bhutta 2012, Bostic and Breck 2003, Dahl et al. 2000). Our counterfactual examines an extreme version of CRA, but this stark benchmark allows us to illustrate the largest potential consequences CRA would have in our model.

Introducing rejections changes the the qualitative and quantitative implications relative to removing rejections altogether. Mortgage rates decrease on average, but only by 28.9bp. The difference is that high types see 9bp increases in mortgage rates on average, while low types see declines of 43bp. In other words, introducing rejections, even if they are uninformative, pushes the predictions of the model closer to the standard pooling intuition, in which bad types gain and good types lose. The mechanism, however, arises through search. Rejections increase de facto search cost relative to no-rejection case, which puts less pressure on banks to lower rates. Nevertheless, offer rates do decline, which would benefit high types, if their search behavior did not adjust. Random rejections of high type borrowers, however, change their de facto search costs increase relative to the estimated model. The resulting change in search behavior of high type borrowers leads them to accept higher rates despite a decline in the rates lenders offer.

As a result, bank profits fall by 6.3bp at the mean realized interest rate. Given a total market size of $479billion, this 6.3bp reduction in profits implies that informative screening is worth approximately $1.2billion (4 * 0.00063 * $479billion) per year. This is substantially less than the no-rejection benchmark. In other words, banks are collectively better off randomly rejecting borrowers, even if these rejections are completely uninformative, since rejections increase borrowers’ willingness to accept higher rate mortgages.

The implications of these results are twofold. First, even though it is somewhat extreme, the counterfactual suggests that it is difficult to implement CRA style policies without decreasing bank profits. What is interesting in our setting is that a large part of the equilibrium response is driven by search responses of borrowers, which puts pressure on offered rates.

\footnote{In order to maintain the same overall application acceptance probability as is observed in our data, we set }
Second, the counterfactuals suggest that improvement in lenders’ screening technology are profitable for lenders, and can result in losses to even high quality borrowers. The idea that low quality borrowers’ search behavior changes the distribution of offer rates, which spills over to high quality borrowers has not been investigated within this setting.

8.3 Tighter Lending Standards

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 at the peak of the crisis during his tenure as chairman of the Federal Reserve.31 Our model provides a unique opportunity to understand the implications of this tighter lending standards along new and crucial dimensions: mortgage pricing and borrowers’ search response. As we show, tightening of lending standards results in higher mortgage rates even if the underlying costs of providing mortgages do not change.

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

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

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

---

32Specifically, we estimate a logit with state and period fixed effects for one of three periods: pre-crisis, during the crisis, and post crisis.
33For a 30-year fixed rate mortgage with principal of $170,000 and interest rate of 4% per year.
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 demand side effects. Indeed, absent these equilibrium side effects, mean rates in the market increase by only 0.8 basis points. Overall, our model illustrates that tighter lending standards result in increased mortgage rates on the order of a discrete increment in the Fed’s policy rate, as well as larger distributional consequences of search costs. A simple way to think of this is that tighter lending standards act as though credit supply is more constrained. Policies affecting credit standards must therefore take into account this direct effect on realized prices in credit markets, in addition to the standard credit access considerations.

8.4 Discrimination and Redlining

Redlining is a practice of discrimination that denies access to products to consumers due to their socioeconomic, racial, or ethnic makeup. Substantial regulation of the mortgage market, including the The Home Mortgage Disclosure Act (HMDA) arose in response to redlining concerns, which are an active area of enforcement and policy.\textsuperscript{34} The search frictions we introduce here permit the analysis of a realistic redlining policy, in which a portion of lenders in the market discriminate by lowering approval rates for borrowers from the discriminated group. Such discrimination is more subtle, and differs from explicitly denying credit to the discriminated group, or charging different prices. Incorporating realistic institutional features of the market for mortgage finance permits us to study the effect of this type of redlining on the discriminated group, as well as the consequences for the equilibrium distribution of interest rates and adverse selection in this market.\textsuperscript{36}

We begin our redlining counterfactual by defining the discriminated group of borrowers, redlining banks, and the nature of redlining discrimination. Potential borrowers belong either to the non-discriminated group $W$, or the discriminated group $B$, the latter comprising 20% of the pool. For expositional clarity, these borrowers have identical search and creditworthiness distributions.\textsuperscript{37} A redlining bank approves the discriminated $B$ borrowers with 50% of the probability that the non-discriminated $W$ borrowers of the same creditworthiness are accepted: that is, $p_B^z = 0.5p_W^z$. Half of banks in the market redline. The non-redlining banks ignore the $B,W$ distinction. Banks can only discriminate based on acceptance probabilities, and have to offer the same interest rates to the discriminated and non-discriminated groups. Preventing discrimination on prices focuses the mechanism on one type of redlining, and is also mostly consistent with the types of redlining and discrimination which have been alleged in this market.\textsuperscript{38}

\textsuperscript{36}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.
\textsuperscript{37}We therefore rule out statistical discrimination, under which the discriminated characteristic would be indicative of underlying type.
Last, we assume that the borrowers are only aware of the proportion of banks redlining, but not which banks redline. This is consistent with the fact that discriminated borrowers keep applying for loans from banks, which are later alleged to have discriminated. The results are presented in Figure 12, and summarized in Table 7.

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

Becker (1957) argues that discriminating firms lose profits. In our setting, redlined lenders’ profits are mostly hurt though lower volumes than the prices they charge. In order to compensate for their lower market share from discriminatory 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 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.

9 Conclusion

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

---

39Note that for the purposes of this counterfactual analysis, we account for the fact that redlining banks will have 50% of the market share that a non-redlining bank will have of each type of borrower when constructing the equilibrium distribution of offered rates in the market.
with asymmetric information. The model predicts that search behavior is not only related to consumer sophistication, as predicted by standard search models, but also by the underlying distribution of types. The interaction between screening and search can explain why borrowers who search a lot obtain expensive mortgages, as well as account for other empirical features of the market, such as the relationship between mortgage approval and search, which standard search models cannot explain. Accounting for the credit approval process is therefore critical in understanding search behavior and equilibrium outcomes in markets for credit products, and more broadly, products in which the seller’s payoff depends on buyer’s characteristics, such as insurance.

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

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

References


### Table 1: Summary Statistics for Mortgages and Applications

<table>
<thead>
<tr>
<th>Search and Rates</th>
<th>Loan Data</th>
<th>Application Data</th>
</tr>
</thead>
<tbody>
<tr>
<td># Inquiries</td>
<td>2.61</td>
<td>12.28</td>
</tr>
<tr>
<td>SD</td>
<td>2.00</td>
<td>11.41</td>
</tr>
<tr>
<td>(Pr{\text{Approval}}) (%)</td>
<td>-</td>
<td>82.19</td>
</tr>
<tr>
<td>Origination Interest Rate (%)</td>
<td>5.69</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>0.86</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Creditworthiness</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FICO</td>
<td>725.80</td>
<td>707.41</td>
</tr>
<tr>
<td>SD</td>
<td>62.52</td>
<td>71.60</td>
</tr>
<tr>
<td>CLTV</td>
<td>73.83</td>
<td>75.34</td>
</tr>
<tr>
<td>SD</td>
<td>18.36</td>
<td>18.31</td>
</tr>
<tr>
<td>Back-end DTI ratio</td>
<td>37.63</td>
<td>37.32</td>
</tr>
<tr>
<td>SD</td>
<td>12.80</td>
<td>12.88</td>
</tr>
<tr>
<td>(Pr{\text{Default}}) (%)</td>
<td>15.95</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>36.61</td>
<td>-</td>
</tr>
<tr>
<td>(Pr{90+\text{ Days Delinquent}}) (%)</td>
<td>9.65</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>29.53</td>
<td>-</td>
</tr>
<tr>
<td>(Pr{\text{Prepay}}) (%)</td>
<td>50.18</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>50.00</td>
<td>-</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Loan Characteristics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>FRM 30-year (%)</td>
<td>76.60</td>
<td>17.86</td>
</tr>
<tr>
<td>SD</td>
<td>42.33</td>
<td>38.30</td>
</tr>
<tr>
<td>FRM 15-year (%)</td>
<td>19.64</td>
<td>8.04</td>
</tr>
<tr>
<td>SD</td>
<td>39.72</td>
<td>27.19</td>
</tr>
<tr>
<td>ARM (%)</td>
<td>3.76</td>
<td>74.10</td>
</tr>
<tr>
<td>SD</td>
<td>19.02</td>
<td>43.81</td>
</tr>
<tr>
<td>Loan Origination Amount ($ 000s)</td>
<td>169.42</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>100.84</td>
<td>-</td>
</tr>
<tr>
<td>Cash-out refi (%)</td>
<td>31.05</td>
<td>0.33</td>
</tr>
<tr>
<td>SD</td>
<td>46.27</td>
<td>5.72</td>
</tr>
<tr>
<td>Rate-term refi (%)</td>
<td>28.13</td>
<td>0.00</td>
</tr>
<tr>
<td>SD</td>
<td>44.96</td>
<td>0.00</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Borrower Characteristics</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>White (%)</td>
<td>79.16</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>40.61</td>
<td>-</td>
</tr>
<tr>
<td>Black (%)</td>
<td>7.52</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>26.38</td>
<td>-</td>
</tr>
<tr>
<td>Borrower Male (%)</td>
<td>42.95</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>49.50</td>
<td>-</td>
</tr>
<tr>
<td>Borrower Age</td>
<td>44.44</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>12.52</td>
<td>-</td>
</tr>
<tr>
<td>Less than High School (%)</td>
<td>25.87</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>43.79</td>
<td>-</td>
</tr>
<tr>
<td>High School and Some College (%)</td>
<td>50.84</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>49.99</td>
<td>-</td>
</tr>
<tr>
<td>College or more (%)</td>
<td>18.35</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>38.71</td>
<td>-</td>
</tr>
<tr>
<td>Borrower Monthly Income</td>
<td>6104.11</td>
<td>-</td>
</tr>
<tr>
<td>SD</td>
<td>6801.63</td>
<td>-</td>
</tr>
<tr>
<td>Investor (%)</td>
<td>8.53</td>
<td>8.06</td>
</tr>
<tr>
<td>SD</td>
<td>27.93</td>
<td>27.23</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Origination Date</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Pre-2006q4 (%)</td>
<td>43.58</td>
<td>48.69</td>
</tr>
<tr>
<td>SD</td>
<td>49.59</td>
<td>49.98</td>
</tr>
<tr>
<td>2006q4-2009q4 (%)</td>
<td>41.68</td>
<td>26.59</td>
</tr>
<tr>
<td>SD</td>
<td>49.30</td>
<td>44.18</td>
</tr>
<tr>
<td>Post-2009q4 (%)</td>
<td>14.74</td>
<td>24.72</td>
</tr>
<tr>
<td>SD</td>
<td>35.45</td>
<td>43.14</td>
</tr>
</tbody>
</table>

| 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

<table>
<thead>
<tr>
<th>Origination Date relative to 2006q4–2009q4:</th>
<th>Loan Data</th>
<th>Application Data</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Pre</td>
<td>During</td>
</tr>
<tr>
<td>Search and Rates</td>
<td></td>
<td></td>
</tr>
<tr>
<td># Inquiries</td>
<td>1.87</td>
<td>3.16</td>
</tr>
<tr>
<td>Pr{Approval} (%)</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Origination Interest Rate (%)</td>
<td>5.91</td>
<td>5.87</td>
</tr>
<tr>
<td>Creditworthiness</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICO</td>
<td>713.00</td>
<td>726.33</td>
</tr>
<tr>
<td>CLTV</td>
<td>74.05</td>
<td>75.46</td>
</tr>
<tr>
<td>Back-end DTI ratio</td>
<td>36.58</td>
<td>39.63</td>
</tr>
<tr>
<td>Pr{Default} (%)</td>
<td>18.32</td>
<td>18.64</td>
</tr>
<tr>
<td>Pr{90+ Days Delinquent} (%)</td>
<td>11.56</td>
<td>10.94</td>
</tr>
<tr>
<td>Pr{Prepay} (%)</td>
<td>46.98</td>
<td>54.25</td>
</tr>
<tr>
<td>Loan Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FRM 30-year (%)</td>
<td>71.68</td>
<td>85.46</td>
</tr>
<tr>
<td>FRM 15-year (%)</td>
<td>23.53</td>
<td>12.72</td>
</tr>
<tr>
<td>ARM (%)</td>
<td>4.78</td>
<td>1.83</td>
</tr>
<tr>
<td>Loan Origination Amount ($ 000s)</td>
<td>138.37</td>
<td>187.32</td>
</tr>
<tr>
<td>Cash-out refi (%)</td>
<td>33.73</td>
<td>30.21</td>
</tr>
<tr>
<td>Rate-term refi (%)</td>
<td>26.69</td>
<td>25.03</td>
</tr>
<tr>
<td>Borrower Characteristics</td>
<td></td>
<td></td>
</tr>
<tr>
<td>White (%)</td>
<td>80.42</td>
<td>77.44</td>
</tr>
<tr>
<td>Black (%)</td>
<td>8.53</td>
<td>8.09</td>
</tr>
<tr>
<td>Borrower Male (%)</td>
<td>44.48</td>
<td>41.95</td>
</tr>
<tr>
<td>Borrower Age (%)</td>
<td>43.55</td>
<td>44.33</td>
</tr>
<tr>
<td>Less than High School (%)</td>
<td>26.36</td>
<td>27.94</td>
</tr>
<tr>
<td>High School and Some College (%)</td>
<td>46.49</td>
<td>53.82</td>
</tr>
<tr>
<td>College or more (%)</td>
<td>16.13</td>
<td>17.94</td>
</tr>
<tr>
<td>Borrower Monthly Income</td>
<td>5087.62</td>
<td>6462.76</td>
</tr>
<tr>
<td>Investor (%)</td>
<td>7.22</td>
<td>9.05</td>
</tr>
<tr>
<td>Observations</td>
<td>573,891</td>
<td>548,819</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th></th>
<th># Inquiries</th>
<th>1st Quartile</th>
<th>2nd Quartile</th>
<th>3rd Quartile</th>
<th>4th Quartile</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Panel A: Mortgage Applicants</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FICO score (standardized)</td>
<td>-3.881***</td>
<td>9.256***</td>
<td>4.831***</td>
<td>-1.345***</td>
<td>-12.743***</td>
</tr>
<tr>
<td></td>
<td>(0.091)</td>
<td>(0.178)</td>
<td>(0.289)</td>
<td>(0.229)</td>
<td>(0.332)</td>
</tr>
<tr>
<td>Combined LTV (Standardized)</td>
<td>0.838***</td>
<td>-2.853***</td>
<td>-0.742***</td>
<td>0.909***</td>
<td>2.686***</td>
</tr>
<tr>
<td></td>
<td>(0.063)</td>
<td>(0.078)</td>
<td>(0.149)</td>
<td>(0.058)</td>
<td>(0.231)</td>
</tr>
<tr>
<td>Back-end DTI Ratio (Standardized)</td>
<td>0.555***</td>
<td>-2.255***</td>
<td>-0.474***</td>
<td>0.914***</td>
<td>1.815***</td>
</tr>
<tr>
<td></td>
<td>(0.018)</td>
<td>(0.133)</td>
<td>(0.086)</td>
<td>(0.084)</td>
<td>(0.072)</td>
</tr>
<tr>
<td>FRM 15-year</td>
<td>1.404***</td>
<td>-4.038***</td>
<td>-1.608***</td>
<td>0.932***</td>
<td>4.714***</td>
</tr>
<tr>
<td></td>
<td>(0.058)</td>
<td>(0.194)</td>
<td>(0.178)</td>
<td>(0.153)</td>
<td>(0.202)</td>
</tr>
<tr>
<td>FRM 30-year</td>
<td>-0.45***</td>
<td>2.266***</td>
<td>0.342*</td>
<td>-1.08***</td>
<td>-1.60***</td>
</tr>
<tr>
<td></td>
<td>(0.035)</td>
<td>(0.252)</td>
<td>(0.187)</td>
<td>(0.158)</td>
<td>(0.128)</td>
</tr>
<tr>
<td>Cash-out refi</td>
<td>-1.045***</td>
<td>1.099*</td>
<td>2.043***</td>
<td>0.682 -3.825</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(0.132)</td>
<td>(0.593)</td>
<td>(0.407)</td>
<td>(0.444)</td>
<td>(0.444)</td>
</tr>
<tr>
<td></td>
<td>(0.142)</td>
<td>(0.370)</td>
<td>(0.364)</td>
<td>(0.379)</td>
<td>(0.458)</td>
</tr>
<tr>
<td>Observations</td>
<td>5202721</td>
<td>5202721</td>
<td>5202721</td>
<td>5202721</td>
<td>5202721</td>
</tr>
<tr>
<td>R²</td>
<td>0.2096</td>
<td>0.1106</td>
<td>0.0190</td>
<td>0.0089</td>
<td>0.1558</td>
</tr>
</tbody>
</table>

| **Panel B: Realized Mortgage Borrowers** |             |              |              |              |              |
| 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      |
| R²                      | 0.2378      | 0.2232       | 0.0100       | 0.0260       | 0.1731       |

Notes: Estimated coefficients from regression equation 1 reported. Panel A reports estimates for the sample of mortgage applications, while Panel B reports estimates for the sample of realized mortgage borrowers. Column 1 reports coefficients from a regression in which the dependent variable is the number of inquiries on an applicant’s credit report. Columns 2 through 5 report coefficients from a regression in which the dependent variable is an indicator variable, scaled by 100, for whether the applicant was in the first, second, third, or fourth quartile of inquiries, respectively. Standard errors clustered at the origination quarter x state level reported in parentheses beneath coefficient. All regressions include origination quarter x state fixed effects. Coefficients marked with *, **, and *** are statistically different from 0 at the 10%, 5%, and 1% level, respectively.
Table 4: Relationship between search and origination rates within demographic groups

<table>
<thead>
<tr>
<th>Years of Education:</th>
<th>Monthly Income:</th>
<th># Inquiries</th>
<th>12</th>
<th>13 - 15</th>
<th>16+</th>
<th>≤ $3,000</th>
<th>$3,001-$7,500</th>
<th>&gt; $7,500</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ 12</td>
<td></td>
<td>-0.009**</td>
<td>-0.012***</td>
<td>-0.010***</td>
<td>-0.004</td>
<td>-0.008***</td>
<td>-0.009***</td>
<td></td>
</tr>
<tr>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13 - 15</td>
<td></td>
<td>0.146***</td>
<td>0.181***</td>
<td>0.146***</td>
<td>0.110***</td>
<td>0.162***</td>
<td>0.161***</td>
<td></td>
</tr>
<tr>
<td>(0.037)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16+</td>
<td></td>
<td>-0.009</td>
<td>-0.002</td>
<td>-0.010**</td>
<td>-0.008**</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>≤ $3,000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$3,001-$7,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.004)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&gt; $7,500</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.003)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| Observations       | 327583         | 652322      | 237401 | 252882 | 748080 | 279436    |
| R²                 | 0.7755         | 0.8094      | 0.8317 | 0.7478 | 0.7970 | 0.8411    |

<table>
<thead>
<tr>
<th>White</th>
<th>Black</th>
<th>Hispanic</th>
<th>Asian</th>
</tr>
</thead>
<tbody>
<tr>
<td># Inquiries</td>
<td>-0.001**</td>
<td>-0.002</td>
<td>-0.010**</td>
</tr>
<tr>
<td>(0.03)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># Inquiries²</td>
<td>0.158***</td>
<td>0.105**</td>
<td>0.166***</td>
</tr>
<tr>
<td>(0.031)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Observations</td>
<td>847288</td>
<td>77009</td>
<td>79678</td>
</tr>
<tr>
<td>R²</td>
<td>0.8078</td>
<td>0.7130</td>
<td>0.7550</td>
</tr>
</tbody>
</table>

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

<table>
<thead>
<tr>
<th>λ</th>
<th>p_h</th>
<th>p_l</th>
<th>p_h - p_l</th>
<th>x_h</th>
<th>x_l</th>
<th>x_h - x_l</th>
<th>μ_c</th>
<th>σ_c</th>
<th>μ_H</th>
<th>σ_H</th>
<th>m</th>
<th>σ_ξ</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.268</td>
<td>1.000</td>
<td>0.193</td>
<td>0.807</td>
<td>1.000</td>
<td>0.410</td>
<td>0.590</td>
<td>-1.284</td>
<td>0.381</td>
<td>0.142</td>
<td>0.547</td>
<td>-1.576</td>
<td>0.410</td>
</tr>
<tr>
<td>(0.002)</td>
<td>(0.004)</td>
<td>(0.003)</td>
<td>(0.003)</td>
<td>(0.001)</td>
<td>(0.004)</td>
<td>(0.005)</td>
<td>(0.004)</td>
<td>(0.002)</td>
<td>(0.001)</td>
<td>-</td>
<td>-</td>
<td></td>
</tr>
</tbody>
</table>

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 14.3.
<table>
<thead>
<tr>
<th></th>
<th>All Borrowers</th>
<th>High Type</th>
<th>Low Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>S.D.</td>
<td>Average</td>
</tr>
<tr>
<td><strong>Realized Interest Rates</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline MLE Estimates</td>
<td>0.027</td>
<td>0.515</td>
<td>-0.228</td>
</tr>
<tr>
<td>Baseline Equilibrium</td>
<td>-0.002</td>
<td>0.664</td>
<td>-0.384</td>
</tr>
<tr>
<td>No Rejection</td>
<td>-0.227</td>
<td>0.370</td>
<td>-0.228</td>
</tr>
<tr>
<td>No Rejection (Eqm.)</td>
<td>-2.101</td>
<td>0.710</td>
<td>-2.103</td>
</tr>
<tr>
<td>Tighter Lending Standards</td>
<td>0.035</td>
<td>0.523</td>
<td>-0.228</td>
</tr>
<tr>
<td>Tighter Lending Standards (Eqm.)</td>
<td>0.252</td>
<td>0.754</td>
<td>-0.217</td>
</tr>
<tr>
<td>Uninformative Screening - CRA</td>
<td>0.007</td>
<td>0.463</td>
<td>0.005</td>
</tr>
<tr>
<td>Uninformative Screening - CRA (Eqm.)</td>
<td>-0.291</td>
<td>0.471</td>
<td>-0.293</td>
</tr>
<tr>
<td>Redlining</td>
<td>0.040</td>
<td>0.517</td>
<td>-0.189</td>
</tr>
<tr>
<td>Redlining (Eqm.)</td>
<td>0.285</td>
<td>0.724</td>
<td>0.243</td>
</tr>
<tr>
<td><strong>Search Distribution</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline Model Estimates</td>
<td>3.40</td>
<td>2.66</td>
<td>1.79</td>
</tr>
<tr>
<td>Baseline Equilibrium</td>
<td>3.53</td>
<td>2.67</td>
<td>2.10</td>
</tr>
<tr>
<td>No Rejection</td>
<td>1.79</td>
<td>1.29</td>
<td>1.79</td>
</tr>
<tr>
<td>No Rejection (w/ Supply)</td>
<td>2.88</td>
<td>2.21</td>
<td>2.88</td>
</tr>
<tr>
<td>Tighter Lending Standards</td>
<td>3.60</td>
<td>2.79</td>
<td>1.79</td>
</tr>
<tr>
<td>Tighter Lending Standards (Eqm.)</td>
<td>3.77</td>
<td>2.80</td>
<td>2.25</td>
</tr>
<tr>
<td>Uninformative Screening - CRA</td>
<td>2.76</td>
<td>2.08</td>
<td>2.76</td>
</tr>
<tr>
<td>Uninformative Screening - CRA (Eqm.)</td>
<td>2.78</td>
<td>2.09</td>
<td>2.78</td>
</tr>
<tr>
<td>Redlining</td>
<td>3.45</td>
<td>2.70</td>
<td>1.77</td>
</tr>
<tr>
<td>Redlining (Eqm.)</td>
<td>3.24</td>
<td>2.74</td>
<td>1.12</td>
</tr>
<tr>
<td><strong>Supply Effects</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Baseline MLE Estimates</td>
<td>0.142</td>
<td>0.547</td>
<td>1.873</td>
</tr>
<tr>
<td>Baseline Equilibrium</td>
<td>0.206</td>
<td>0.723</td>
<td>1.893</td>
</tr>
<tr>
<td>No Rejection</td>
<td>-0.732</td>
<td>1.259</td>
<td>0.001</td>
</tr>
<tr>
<td>Tighter Lending Standards</td>
<td>0.142</td>
<td>0.547</td>
<td>1.860</td>
</tr>
<tr>
<td>Tighter Lending Standards (Eqm.)</td>
<td>0.483</td>
<td>0.805</td>
<td>2.130</td>
</tr>
<tr>
<td>Uninformative Screening - CRA</td>
<td>0.142</td>
<td>1.873</td>
<td>1.867</td>
</tr>
<tr>
<td>Uninformative Screening - CRA (Eqm.)</td>
<td>-0.147</td>
<td>0.561</td>
<td>1.830</td>
</tr>
<tr>
<td>Redlining</td>
<td>0.142</td>
<td>0.547</td>
<td>1.860</td>
</tr>
<tr>
<td>Redlining (Eqm.)</td>
<td>0.300</td>
<td>0.733</td>
<td>1.995</td>
</tr>
</tbody>
</table>

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 TIEV profit shocks $\xi_{jk}$. 
Table 7: Redlining Counterfactual Summary

<table>
<thead>
<tr>
<th>Borrower Information</th>
<th>Realized Rates</th>
<th>Search</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>S.D.</td>
<td>Average</td>
</tr>
<tr>
<td>Redlined Group</td>
<td>0.298</td>
<td>0.723</td>
<td>3.54</td>
</tr>
<tr>
<td>Non-Redlined Group</td>
<td>0.282</td>
<td>0.725</td>
<td>3.17</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Bank Information</th>
<th>Offered Rates</th>
<th>Expected Profits</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Average</td>
<td>S.D.</td>
</tr>
<tr>
<td>Redlining Banks</td>
<td>0.291</td>
<td>0.753</td>
</tr>
<tr>
<td>Non-Redlining Banks</td>
<td>0.308</td>
<td>0.713</td>
</tr>
</tbody>
</table>

Notes: Table reports mean and standard deviation of search and realized interest rates for the group of redlined borrowers R 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 report the mean and standard deviation of the offered rate distribution for redlining and non-redlining banks, estimated according to the routine outlined in Section 15.2. 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 TIEV profit shocks $\xi_{jk}$. 
Figure 1: Distribution of Mortgage Rates in the U.S.

Panel A: Raw Rates by Origination Date

Panel B: Raw Rates by Borrower FICO Score

Panel C: Rates Residualized Against Observables

Notes: Figure plots the kernel-density estimated distribution of mortgage rates in the U.S. Panel A plots the raw observed rates across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post-crisis period from the first quarter of 2010 on. Panel B plots the distribution of observed mortgage rates for three borrower FICO buckets: low FICO (≤ 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 × origination quarter fixed effects.
Figure 2: Inquiry distribution among mortgage applicants

Panel A: Loan-Level Dataset

Panel B: Applicant Dataset

Panel C: Applicants by FICO Score

Panel D: Applicants by Education

Notes: Figure plots the distribution of inquiries across successful mortgage applicants (i.e., those in our loan-level dataset) across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post crisis period from the first quarter of 2010 on. Dashed lines plot bootstrapped 95% confidence intervals, in which bootstraps are clustered at the origination quarter level. Panel A plots the inquiry distribution for all borrowers in our application-level dataset, while Panel B plots the inquiry distribution for our loan-level dataset containing borrowers with successful application. Panel C plots the distribution of inquiries across mortgage applicants for three FICO buckets: low FICO ($\leq 620$), middle FICO (620-719) and high FICO (720+). Panel D plots the distribution of inquiries across successful mortgage applicants (i.e., those in our loan-level dataset) for three borrower education groups.
Figure 3: Rates and search by FICO bucket

Panel A: All Borrowers

Panel B: FICO \leq 620

Panel C: 620 < FICO \leq 720

Panel D: FICO > 720

Notes: Figure plots average realized interest rates against inquiry counts for realized loans for all borrowers and across three FICO buckets: low FICO (\leq 620), middle FICO (620-719) and high FICO (720+).
Figure 4: Relationship Between Search and Mortgage Origination Rates, Conditional on Observables

Panel A: All Borrowers

Panel B: FICO ≤ 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 5: Characteristics of a Sequential Search Model with Informative Screening

Panel A: Reservation rates of high and low types

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

Panel C: Borrower type distribution and search

Panel D: Relationship between search and prices

Panel E: Relationship between search and default rate

Panel F: Relationship between search and application approval

Notes: Figure plots key aspects of the mortgage market under the baseline model with informative screening. Data are simulated from a model in which application approval parameters are set to \( p_h = 0.95 \) and \( p_l = 0.05 \), the share of high types is \( \lambda = 0.7 \), the probability of full repayment for high and low types are \( x_h = 0.8 \), and \( x_l = 0.4 \), respectively, and the search costs and offered rates are distributed according to truncated normal distributions. Panel A plots the distribution of reservation rates for high type (in blue) and low type (in red) borrowers. Panel B plots the percent of borrowers that are high type at each realized interest rate, highlighting the pattern of adverse selection when screening is present. Panel C shows the percentage of successful borrowers who are high type as a function of search. Panel D, E, and F display the relationship between search and realized interest rates, eventual mortgage default rate, and application approval probability, respectively.
Notes: Figure plots average default rates against search. Panel A defines default to be serious (90+ days) delinquency, or foreclosure, while Panel B limits attention to seriously delinquent loans. Panels C through F plot regression coefficients estimated from equation 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, back end 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 7: Relationship Between Search and Mortgage Application Approval Rates

Panel A: All Applicants

Panel B: By Borrower FICO Score

Panel C: By Origination Date

Notes: Figure plots the relationship between application approval rate and the number of inquiries on an applicant’s credit report. A line of best fit, weighted by the number of applicants with $s$ inquiries, is drawn as a visual aid. Panel A plots the relationship for all applicants in our application dataset. Panel B displays the relationship for three applicant FICO score buckets separately. The Low FICO group (in red) contains those with FICO score below 620, Mid FICO (in blue) corresponds to those with a FICO score between 620 and 720, while the High FICO group (in green) shows the patterns for those with a FICO score above 720. Panel C shows the patterns across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post crisis period from the first quarter of 2010 on. Lines of best fit, weighted by the number of applicants with $s$ inquiries, drawn as a visual aid.
Figure 8: Relationship between search and mortgage application approval rates, conditional on observables by FICO bucket

Panel A: All Applicants

Panel B: FICO \leq 620

Panel C: 620 < FICO \leq 720

Panel D: FICO > 720

Notes: Figure plots regression coefficients estimated from equation 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, back-end 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.
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.
Panel E: Share of high types as function of origination rate

Notes: Figure plots the performance of our model under our benchmark estimated parameters from Table 5. Black lines plot quantities in our estimation sample, while light blue lines plot those implied by a large model simulation using parameters estimated by maximum likelihood following the approach laid out in section 7.1. Origination rates in data residualized against the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination year fixed effects. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C and D show the relationship between search and origination interest rates and default probability, respectively, where default probability is measured as of January 2015. To compute these default probabilities in the simulation, we randomly draw a mortgage's origination date from the distribution of origination dates in the data. Panel E shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate \( r \) who are of high type.
Figure 11: Tighter Lending Standards Counterfactual

Panel A: Realized Rate Distribution
Panel B: Distribution of Search
Panel C: Relationship between search and price
Panel D: Relationship between search and default
Panel E: Relationship between search and application approval
Panel F: Share of high types as function of origination rate

Notes: Figure plots the key aspects of search behavior under our baseline parameter estimates (black line) and a model in which the odds of application approval drop as they did following the recession (light blue line), allowing the equilibrium offered rate distribution to adjust. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C, D, and E show the relationship between search and origination interest rates, the probability of ever defaulting, and the application approval rate. Panel F shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate $r$ who are of high type.
Figure 12: Redlining Counterfactual

Panel A: Realized Rate Distribution

Panel B: Distribution of Search

Panel C: Relationship between search and price

Panel D: Relationship between search and default

Panel E: Relationship between search and application approval

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

Notes: Figure plots the key aspects of search behavior under our baseline parameter estimates (black line) and 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.
Figure 13: 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.
## Appendix A: Additional Robustness Tables and Figures

Table 8: $k$-means Clustering Test for Multiple Borrower Types

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

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 14: 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 15: Relationship between Search and Realized Mortgage Interest Rates, Conditional on Observables, by Ex-Post Delinquency Status and Brokerage Status

Panel A: Loans which do not default ex post

Panel B: Loans which default ex post

Panel C: Unbrokered loans

Panel D: Brokered loans

Notes: Figure plots regression coefficients estimated from equation 2 using OLS for the separate subsamples of loans which do not default ex post (Panel A), which do default ex post (Panel B), which are mortgages found without a broker (Panel C), and which are brokered mortgages (Panel D). The dependent variable in each regression is the origination interest rate on a loan. Default defined as a loan being in foreclosure or at least 90 days delinquent by Jan 1, 2015. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to $s$ for $s$ in $\{2, 3, 4, ..., 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower’s FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.
Figure 16: 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 17: Default rate and search: by education and income level

- **Panel A**: High School or Less
- **Panel B**: Some College
- **Panel C**: College Graduate
- **Panel D**: Monthly Income ≤ $3,000
- **Panel E**: $3,000 < Income ≤ $7,500
- **Panel F**: Monthly Income > $7,500

Notes: Figures plots average annualized default rates against search, subsetting borrowers according to their education and monthly income levels.
Figure 18: Default rate and search: by borrower race

Panel A: White
Panel B: Black
Panel C: Asian
Panel D: Hispanic
Notes: Figures plots average annualized default rates against search, subsetting borrowers according to their race.
Figure 19: Relationship Between Search and Realized Interest or Default Rates, Controlling for LLPA Categories

Notes: Figure plots regression coefficients estimated from equation 2 using OLS. The dependent variable in Panels A and B is the origination interest rate on a loan, while for Panels C and D it is an indicator for the loan being in foreclosure or at least 90 days delinquent by Jan 1, 2015. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in \{2, 3, 4, \ldots, 11+\}. The omitted category is s = 1. Controls are included for every bin for loan-level price adjustment as urged by Fannie Mae, available at https://www.fanniemae.com/content/pricing/llpa-matrix.pdf. Panels B and D additionally include state fixed effects, lender fixed effects, and origination quarter fixed effects. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.
Figure 20: Search Behavior of Rarely Rejected Borrowers - Alternative Rarely Rejected Definition

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 applying 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 21: Estimates by subsample

Panel A: Proportion of high types $\lambda$

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 $\sqrt{(e^{\sigma_c^2} - 1) e^{(2\mu_c+\sigma_c^2)}}$

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.
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 8. 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.

Appendix C: Robustness of Positive Relationship Between Search and Rates

Figures 14-15 and 17-19 present robustness of the key empirical fact of the paper, namely that realized interest and default rates increase in borrower search. Figure 14 plots the mean realized interest rate against against search for a host of borrower subsets - by race, education, income, and product type. Figure 15 plots the estimated regression coefficients from equation 2 for the subset of loans that do/do not eventually default, and for the set of borrowers who do/do not obtain their mortgage from a broker. In all cases, we find the positive relationship between search and interest rates. Figures 17 through 18 plot the mean default rates as a function of search by borrower race, and monthly income. Again, we consistently find a positive relationship between search and default. Finally, Figure 19 plots coefficients estimated from Equations 2 (Panels A and B) and 7 (Panels C and D), after increasing our set of controls to include every bucket for loan-level price adjustments provided by Fannie Mae.\footnote{These adjustment factors may be obtained from https://www.fanniemae.com/content/pricing/llpa-matrix.pdf.} Panels A and C show the results when we omit state, origination date, and lender fixed effects, while Panels B and D include our full suite of fixed effects. Without controlling for aggregate trends, the relationship between search and interest rates becomes noisier. However, controlling for borrower state and origination quarter recovers the positive relationship between interest rates, default, and search. This is unsurprising - the unobserved quality of borrowers is thought to have

65
changed substantially over our sample period, and varies substantially across states. Overall, the central fact of the paper appear robust to all manner of control variables and across nearly all subsets of borrowers.

14 Appendix D: Likelihood Construction

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

$$Pr \{ R \leq r, S = s | z, r^* \} = Pr \{ (O_s \leq r \leq r^* \cap A_s = 1) \cap \neg(O_1 \leq r^* \cap A_1 = 1) \cap \ldots \cap \neg(O_{s-1} \leq r^* \cap A_{s-1} = 1) \cap z, r^* \}$$

$$= Pr \{ (O_s \leq r \leq r^* \cap A_s = 1) | z, r^* \} (Pr \{ \neg(O \leq r^* \cap A = 1) | z, r^* \})^{s-1}$$

$$= 1 \{ r \leq r^* \} \cdot p_z h(r) (1 - p_z H(r^*))^{s-1}$$

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:

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

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.
\[ l(R_i = r, S_i = s | A_{is} = 1, \chi_{is} = 1) = \lambda p_h(r) \int_r^\infty (1 - p_h H(r^*))^{s-1} dF_h(r^*) \]
\[ + (1 - \lambda) p_h(r) \int_r^\infty (1 - p_h H(r^*))^{s-1} dF_i(r^*) \]

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:

\[ l(R_i = r, S_i = s, D_i = d | A_{is} = 1, \chi_{is} = 1, t, T) = \lambda \left( d(1 - x_h^{A_{is}/T}) + (1 - d)x_h^{A_{is}/T} \right) p_h(r) \int_r^\infty (1 - p_h H(r^*))^{s-1} dF_h(r^*) \]
\[ + \left(1 - \lambda \right) \left( d(1 - x_i^{A_{is}/T}) + (1 - d)x_i^{A_{is}/T} \right) p_h(r) \int_r^\infty (1 - p_i H(r^*))^{s-1} dF_i(r^*) \]

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

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

It is straightforward to show that the numerator may be written as
\[
Pr\{s - 1 \text{ failed inquiries} \cap \chi_{is} = 1\} = \lambda \int H(r^*) (1 - p_h H(r^*))^{s-1} dF_h(r^*) \\
+ (1 - \lambda) \int H(r^*) (1 - p_l H(r^*))^{s-1} dF_l(r^*)
\] (10)

It remains to derive \(Pr\{\chi_{is} = 1\}\), 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 yet to originate a mortgage, the probability that we observe inquiry \(s'\) is then

\[
\frac{1}{\tilde{S}} Pr\{s' - 1 \text{ failed inquiries} \cap \chi_{is'} = 1\} = \frac{1}{\tilde{S}} \lambda \int H(r^*) (1 - p_h H(r^*))^{s'-1} dF_h(r^*) \\
+ \frac{1}{\tilde{S}} (1 - \lambda) \int H(r^*) (1 - p_l H(r^*))^{s'-1} dF_l(r^*)
\]

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 \(\tilde{S}\):

\[
Pr\{s - 1 \text{ failed inquiries} | \chi_{is} = 1\} = \frac{Pr\{s - 1 \text{ failed inquiries} \cap \chi_{is} = 1\}}{\sum_{1 \leq s' \leq \tilde{S}} Pr\{s' - 1 \text{ failed inquiries} \cap \chi_{is'} = 1\}}
\] (11)

The denominator may be written as

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

Letting \(\tilde{S}\) go to infinity and substituting back into 11 yields the likelihood contribution of an application with \(s\) inquiries:

\[
l(S_i = s|\chi_{is} = 1) = \frac{Pr\{s - 1 \text{ failed inquiries} \cap \chi_{is} = 1\}}{\lambda/p_h + (1 - \lambda)/p_l}
\] (12)

where the numerator is defined as in equation 10. Combining this with the likelihood of each realized loan from equation 9 yields the likelihood for our full data.\(^{41}\)

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

\(^{41}\)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.
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 $r^*$

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

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

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

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

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

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 $\beta_H \equiv \{\mu_H^{(n)}, \sigma_H^{(n)}, \pi_H^{(n)}\}_{n=1}^N$, while the search cost distribution is well-approximated by a mixture of $N$ log-normally distributed random variables parameterized by $\beta_G \equiv \{\mu_G^{(n)}, \sigma_G^{(n)}, \pi_G^{(n)}\}_{n=1}^N$. That is, we assume that we may write

$$h(r) \approx \sum_n \pi_H^{(n)} \frac{1}{\sigma_H^{(n)} \sqrt{2\pi}} \exp \left[ -\frac{(r - \mu_H^{(n)})^2}{2 \sigma_H^{(n)}^2} \right]$$

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

for $\pi^{(n)}$ the mixing weight on the $n^{th}$ normal distribution, $\mu^{(n)}, \sigma^{(n)}$ the mean and standard deviation parameters of the $n^{th}$ underlying normal distribution. This assumption permits the analytical construction of the reservation rate distribution for high and low type individuals, and is motivated by the roughly normal distribution of residualized realized rates observed in Figure 1. A detailed description of this construction is provided in Appendix 15.1.

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 \( \mathcal{L} \), while the set of observations in the application dataset be given by \( \mathcal{A} \). We therefore maximize the following log-likelihood with respect to a choice of \( \theta \equiv \{ p_h, p_l, x_h, x_l, \lambda, \beta_H, \beta_C \} \)

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

where \( l(R_i, D_i, S_i|A_{is} = 1, \chi_{is} = 1, t, T) \) is given by equation 9, and \( l(S_i = s|\chi_{is} = 1, \theta) \) is given by equation 12. Since \( q \) is additively separable from \( \theta \), its value will not affect our optimal choice of \( \hat{\theta} \). To uniquely identify the parameters, we impose that \( p_h \geq p_l \), but impose nothing about the relationship between \( x_h \) and \( x_l \).

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

### 14.2 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, reflecting the market power that the search process gives banks:

$$\frac{dq_z(r)}{dr} = -\frac{f_z(r)}{H(r)} < 0.$$  \hfill (13)

### 14.3 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 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 $\hat{x}_z = (x_z - 1)/\log(x_z)$ of every dollar loaned to a type $z$ borrower.\footnote{To see this, suppose a borrower originates a mortgage whose term is $T$, requiring $N$ discrete payments of equal size. Letting $\Omega(t)$ be the survival function after a fraction $t$ of the loan's life, we have that the expected repayment is $\sum_{1 \leq n \leq N} \Omega(nT/N)/N$. Substituting in for $\Omega(t)$ using the proportional hazard assumption implies that the expected repayment can be expressed as

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

As a result, letting $S$ be the size of the market, the expected profits from making a loan at rate $r$ are

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

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

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

15 Appendix E: Computational Details

15.1 Constructing Reservation Rate Distributions from Search Cost Distributions

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

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

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

$$h(r) \approx \sum_n \pi^H_n \frac{1}{\sqrt{2\pi} \sigma^H_n} \exp \left[ -\frac{(r - \mu^H_n)^2}{2 (\sigma^H_n)^2} \right]$$

for $\pi^H_n$ the mixing weight on the $n$th normal distribution, $\mu^H_n, \sigma^H_n$ the mean and standard deviation parameters of the $n$th underlying normal distribution. Similarly, we assume that the search cost distribution is well approximated by a mixture of log-normal distributions parameterized by $\beta_G \equiv \{\mu^G_n, \sigma^G_n, \pi^G_n\}_{n=1}^N$.

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

$$\psi(r^*) = p_z \int_{-\infty}^{r^*} (r^* - r) \, dH(r)$$

$$= p_z r^* H(r^*) - p_z \sum_n \pi_n \int_{-\infty}^{r^*} \frac{r}{\sigma_n \sqrt{2\pi}} \exp \left[ -\frac{(r - \mu_n)^2}{2 (\sigma_n^2)^2} \right] \, dr$$

$$= p_z r^* H(r^*) - p_z \sum_n \pi_n \left[ \mu_n cdf_{N(\mu_n,\sigma_n)}(r^*) - \sigma_n^2 pdf_{N(\mu_n,\sigma_n)}(r^*) \right]$$
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 $H(o)$.

### 15.2 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 (13). Meanwhile, changes to $\lambda, x_h, m,$ 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 (13) 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 minimize the distance between $H(r; \beta_H)$ and $\hat{H}(r; \beta_H)$ with respect to $\beta_H$. The problem may then be written as

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

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.\(^{43}\) 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

\(^{43}\)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.
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 14 therefore compares a normal distribution with logit-implied choice probabilities, which will naturally involve some error. To evaluate the severity of this concern, we search for an equilibrium using the set of parameters estimated using our maximum likelihood routine. The mean and standard deviation of the MLE offered rate distribution are 0.142 and 0.547, respectively. By comparison, the “equilibrium distribution,” obtained by running these parameters through the equilibrium search routine described above has a mean and standard deviation of 0.206 and 0.723, respectively. Although imperfect, we consider this error to be relatively small. After simulating the demand side of the model, this leads to a gap in average rates paid of 2.9bp, and an increase in search of 0.13 inquiries per borrower. For all counterfactuals in which we allow the offered rate distribution to adjust, we compare the counterfactual output against “equilibrium” simulations, which are based on a normally-distributed offered rate distribution with mean and standard deviation of 0.206 and 0.723, respectively.