Submitted to *Econometrica*

SEARCHING FOR APPROVAL

SUMIT AGARWAL
National University of Singapore

JOHN GRIGSBY
Princeton University

ALI HORTAÇSU
University of Chicago

GREGOR MATVOS
Kellogg School of Business

AMIT SERU
Stanford University

VINCENT YAO
Georgia State University

This paper theoretically and empirically studies the interaction of search and application approval in credit markets. Risky borrowers internalize the probability that their application is rejected and behave as if they had high search costs. Thus "overpayment" may be a poor proxy for consumer sophistication since it partly represents rational search in response to rejections. Contrary to standard search

Sumit Agarwal: ushakri@yahoo.com
John Grigsby: john.grigsby@princeton.edu
Ali Hortaçsu: hortacsu@uchicago.edu
Gregor Matvos: gregor.matvos@kellogg.northwestern.edu
Amit Seru: aseru@stanford.edu
Vincent Yao: wyao2@gsu.edu

We thank Matteo Benetton, Peter DeMarzo, Jean-Pierre Dubé, Andreas Fuster, Jean-Francois Houde, Matt Gentzkow, Ralph Koijen, Aviv Nevo, Chris Palmer, Tarun Ramadorai, Nick Roussanov, Tobias Salz, Johannes Stroebel, Adi Sunderam, Motohiro Yogo as well as numerous seminar participants. Sumit Agarwal acknowledges the financial support from the National Natural Science Foundation of China (Grant Nos. 72150004). Kyeongbae Kim and Marco Loseto provided exceptional research assistance.

models, our model implies 1) endogenous adverse selection through the search and application approval process, 2) a possibly non-monotone or non-decreasing relationship between search and realized interest, default, and application approval rates and 3) search costs estimated from transaction prices alone are biased. We find support for the model's predictions using a unique dataset detailing search behavior of mortgage borrowers. Estimating the model, we find that screening is informative and search is costly. Counterfactual analyses reveal that tightening lending standards and discrimination through application rejection both increase equilibrium interest rates. This increase in realized interest rates is in part due to strategic complementarity in bank rate setting.

KEYWORDS: Mortgages, Search, Screening.

1. INTRODUCTION

Search is a leading explanation for price dispersion in consumer credit markets.[1] This paper studies how a key feature of credit markets affects search behavior and pricing: creditors evaluate borrowers' creditworthiness and may reject customers based on this evaluation. If an application is rejected, customers must search for a loan from another lender. Approval is a necessary step to obtain a mortgage, credit card, student, auto or small business loan, and a similar process takes place in the private insurance industry. Survey evidence suggests borrowers' concerns of qualifying for a loan are an important consideration in how they search.[2] We develop a quantitative search model which incorporates the approval process in credit markets. The possibility of rejection alters search behavior and the conclusions that researchers can draw about search from the data. We apply the model to the $2 trillion per year mortgage origination market using a unique proprietary dataset of conforming mortgages from a government sponsored entity (GSE) in the United States.

Our model is an extension of the standard sequential search model with posted prices proposed by Carlson and McAfee (1983) and McCall (1970). We depart from the standard framework by allowing borrowers to differ in their ability to repay the loan and assuming that this creditworthiness is private information. After a credit application is submitted, lenders conduct an in-depth screen of the borrower to obtain an imperfect but informative signal of their creditworthiness. The lender can then either approve a loan or reject the application. If the application is rejected, the borrower must search for another lender, incurring their search cost once more. In the short window during which borrowers are searching, lenders do not observe the past search history and must perform their own in-depth screening.

The approval process alters borrowers' search because the possibility of future rejection increases borrowers' willingness to take a high price offer. An individual borrower whose probability of rejection is $(1-p)$, realizes the benefit of search with probability p. They are therefore willing to accept the same offer as a borrower who is approved for sure, but whose search cost is $\frac{1}{p}$ times higher. This simple intuition has several implications. Since search data is often unavailable, search costs are often estimated by inverting transaction prices to recover

[1] Similar borrowers obtain credit with substantially different interest rates or fees in markets for mortgages(Gurun et al. (2016); Allen et al. (2014); Woodward and Hall (2012); Stroebel (2015)), credit cards (Ausubel (1991); Calem and Mester (1995); Agarwal et al. (2018); Stango and Zinman (2013)), and auto loans (Argyle et al. (2017)).

[2] In the National Survey of Mortgage Originators (NSMO) (28%) of respondents who apply to multiple lenders say they do so out of concern for qualifying for a loan.

the distribution of reservation prices (e.g. Hortaçsu and Syverson (2004); Allen et al. (2014); Roussanov et al. (2018)). This paper argues that applying this approach to credit markets overestimates search costs by $\frac{1}{p}$ because the variation in estimated search cost across consumers confounds variation in search costs and rejection rates.

More broadly, less educated, poor, low credit score (subprime), or minority borrowers' willingness to accept high priced credit products has often been attributed entirely to the lack of their financial sophistication, i.e. to high search costs (Bhutta et al. (2021)). This intuition has been used to target financial sophistication as a way to improve borrower outcomes. Our model suggests that paying high interest rates is a poor proxy for borrower sophistication. These borrowers face a high probability of rejection and are therefore rationally willing to pay high prices. The "financial unsophistication" explanation is especially misleading for minority borrowers who may face higher rejection rates due to discrimination.

In addition, the approval process generates endogenous adverse selection. The possibility of rejection looms larger for less creditworthy borrowers because an in-depth check by the lender is likely to result in application rejection. Consequently, low creditworthiness borrowers behave as if they have higher search costs and are more willing to take up expensive loans. Therefore, lower quality borrowers sort to firms which charge higher prices. This adverse selection arises regardless of whether risky borrowers are more or less willing to pay for credit, which is what drives adverse selection in standard selection models. This result suggests that adverse selection could be endemic in credit and insurance markets, in which lenders screen and reject borrowers.

The model elucidates an important strategic complementarity in bank rate setting. As rejection probabilities rise, borrowers become more willing to accept high interest rate loans. Thus, the residual demand curve facing lenders becomes more inelastic and lenders increase their offered rates in equilibrium. This strategic complementarity and the "cream-skimming" motive from endogenous adverse selection introduce two ways in which lending standards influence bank rate setting. This supply response amplifies the effect of credit standards on realized prices.

The model generates several new testable predictions. First, average transaction prices need not decline monotonically with search. Canonical search models predict that frequent-searchers are mostly those with low search costs who are only willing to accept low prices. Application rejections introduce a second group of frequent-searchers: those who are often rejected. These borrowers are willing to accept high prices and thus sort to high rate lenders because of their search behavior and not directly due to their low creditworthiness. The mixing of these two groups of searchers may thus generate a decreasing, non-monotone or even increasing relationship between search and realized interest rates through the application approval process.

Second, our model predicts a relationship between search, default and loan approvals. Because frequently rejected borrowers are likely of low creditworthiness, frequent-searchers are more likely to default ex post. For the same reason, our model also generates a negative relationship between search and application approval: informative screening reveals frequent-searchers to be creditworthy less often than it does for infrequent-searchers.

We apply the model to study consumer search in the annual $2 trillion mortgage origination market. We use a unique proprietary dataset of conforming mortgages from a large government sponsored entity (GSE) in the United States matched with consumer credit reports from a large national credit bureau. These data allow us to link borrowers' search behavior, proxied by inquiry counts, with mortgage rates, delinquency and application acceptance decisions, conditional on many borrower and loan characteristics.

Consistent with the model of rejection, we find that frequent-searchers obtain higher rate mortgages than borrowers who search little. The relationship between search and transaction prices is rarely tested because search is rarely observed in the data. The fact that mortgage rates, inclusive of all fees, do not decline monotonically with search is very robust and contradicts standard search models without rejection. It survives across different subsamples of borrowers, and after extensive controls for the borrower characteristics that lenders use to set mortgage rates such as FICO, LTV, debt-to-income (DTI), property location and origination date.

The data also show that borrowers who search more are less likely to repay loans ex post, even conditional on observable characteristics. This is the case even though prior borrower searches are not observed by the lender. Linking approval data with search, we show a robust negative relationship between the probability of mortgage approval and the number of searches.

As further validation of the model, we examine a population of borrowers who face almost no possibility of application rejection as a "placebo" test. These borrowers have average rejection rates below 1.25%, which is much lower than in the overall population. In the absence of any possibility of application rejection, our model collapses to the canonical model, predicting a negative relationship between search and realized prices. Strikingly, mortgage origination rates are monotonically decreasing in the frequency of search for the population of "rarely-rejected borrowers" in the data. These results suggest that the non-negative relationship between search and mortgage rates for the overall population is indeed driven by the approval process rather than other unobservable borrower characteristics.

We structurally estimate the model's parameters with maximum likelihood. Consistent with the existing literature on search in mortgage markets, estimated search costs are large: on average, each additional search is equivalent to paying an additional 29.7 basis points (bp) on a loan. For an average loan in our sample, this equates to a cost of $30 per month or about $1,800 over 5 years. In addition, consumers differ in search costs, with a standard deviation of 11.8 bp.

Not accounting for rejection significantly impacts search cost estimates if the estimation is based only on price and quantity data as in Hortaçsu and Syverson (2004). For example, assuming a 25% probability of rejection as in data from the Home Mortgage Disclosure Act (HMDA), estimating search costs in the traditional way inflate search cost estimates by a factor of $\frac{1}{75\%} = 1.33$, or 33%. Standard approaches may estimate search costs to be more than twice their true value for borrowers rejected at a higher rate: those with high debt-to-income and loan-to-value (LTV) ratios. On the other hand, ignoring rejections underestimates search cost if estimation is based on search data in addition to price and quantity data. Doing this in our data underestimates average search costs by a factor of 2, or about 15bp. Intuitively, the model without rejection interprets a large number of searches as a consequence of low search costs, instead of high rejection probability.

We study two counterfactuals. We first consider a situation in which credit availability is limited by tight lending standards even if the costs of funds is low: "mortgages are cheap if you can get one." These situations arose during the financial crisis of 2008 (Mian and Sufi (2009); Stroebel (2015)) and more recently in the beginning of the COVID-19 pandemic.[3] Tighter lending standards increase equilibrium mortgage rates because higher chances of rejection increase borrowers' willingness to accept higher cost mortgages. Lender rejection and rate setting are complements in equilibrium, resulting in substantially higher transaction and posted prices than the borrower responses alone would predict. We estimate that tighter lending standards during the crisis increased average mortgage rates by 25 bp, or half a standard deviation. This large effect in part arises due to the strong strategic complementarity in the lender rate setting decision.

Last, we analyze a realistic redlining policy in which a portion of lenders in the market discriminate by lowering approval rates for borrowers from the discriminated group. Discrimination of this sort was one of the primary reasons for the establishment of the Home Mortgage Disclosure

[3] See, for instance, this article from Mortgage Professional America (MPA) [Accessed on 4/22/2020].

Act (HMDA), a federal law requiring mortgage lenders to submit records of mortgage applications and rejection decisions to regulators. We show that redlining behavior induces borrowers from the discriminated group to pay higher interest rates in equilibrium, even if they purchase a mortgage from a lender that itself does not engage in redlining – non-redlining lenders respond strategically to the inelastic demand of discriminated groups. The search behavior of discriminated groups resembles that of the financially unsophisticated, but in fact reflects the rational incorporation of an increased rejection probability into 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 would increase by 29 bp.

Our paper contributes to the recent literature on price dispersion and choice frictions in the mortgage market (Gurun et al. (2016); Allen et al. (2014); Woodward and Hall (2012); Alexandrov and Koulayev (2017)). Allen et al. (2019) conduct a detailed study of the role played by incumbency advantage and market power on search outcomes in the Canadian mortgage market. Rational inattention has also been proposed as a possible explanation for dispersion in mortgage rates and the low take-up of beneficial refinancing opportunities (Andersen et al. (2015)), and perhaps offers one possible microfoundation for search costs.

Ambokar and Samaee (2019a) find that search costs significantly inhibit refinancing, both directly and by indirectly giving market power to lenders.[4] Ambokar and Samaee focus on the causes and consequences of inaction in the mortgage refinance market using models in which borrowers' creditworthiness is known. Search frictions similarly give lenders market power in our model. Our focus, however, is on the role played by the screening mechanism on search behavior and price setting. We show that ignoring rejection can lead to biased search cost estimates, while screening may result in endemic advese selection. We also develop a framework that allows us to examine the quantitative importance of informative screening and approval in the face of asymmetric information for credit pricing and search behavior.

The role played by switching costs/consumer inertia was studied by Handel et al. (2015) in the context of health insurance. 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 (Lester et al. (2016); Guerrieri et al. (2010)). In our model, borrowers are

[4] Ambokar and Samaee (2019b) show that mortgage search dampens the refinancing channel of monetary policy in a New Keynesian model. This largely operates through lenders' ability to statistically discriminate by charging non-searchers a higher interest rate in the belief that they are thus captive shoppers.

offered only one contract and screening is performed through a noisy technology reflecting the loan approval process. As a result, equilibrium is much easier to compute than in other settings, permitting estimation and counterfactual exercises. While the menu of contracts approach depicts many insurance markets accurately, we believe our model is a more realistic description of the mortgage market and other consumer credit markets.

More broadly, our paper links to the literature using quantitative models to study the effect of competition in financial markets (Benneton (2019); Koijen and Yogo (2016, 2020); Agarwal et al. (2018); Argyle et al. (2017); Buchak et al. (2018, 2020); Gilbukh and Goldsmith-Pinkham (2020); Piazzesi et al. (2020); Scharfstein and Sunderam (2017); Wong (2019)). Our model differs with its focus on the interaction of search and screening. We show that improved screening technology can increase interest rates even for the most creditworthy through spillovers in lenders' rate posting. Our model highlights that search interacts with the rejection and pricing behavior of intermediaries, shaping policy outcomes that occur when lending standards tighten.

We describe the mortgage application process and present our model of search with screening in section 2. Section 3 describes the data used in our empirical analysis. Section 4 tests the model's predictions in the mortgage market. We describe the estimation of our model in section 5 and report its results. Finally, section 6 presents our counterfactual analyses. Section 7 concludes.

2. MODEL

2.1. *Credit Approval Process and Model Setting*

The process obtaining credit — be it a credit card, auto loan, mortgage, other consumer credit or sometimes small business credit — starts with the borrower filing an application, which provides information required by the lender such as the borrower's income, occupation, existing debts and assets. Next, the lender assesses the borrower's creditworthiness by checking their credit report. The credit report of the borrower is "pulled" from a credit bureau.[5] Each pull is recorded as "an inquiry" by the credit bureau. After pulling the credit report, the lender offers the borrower a rate tailored to their characteristics, which the borrower may accept or reject. Thus, a borrower only observes a rate tailored for them *after* their credit is pulled and an inquiry is recorded.[6]

[5]These reports are typically provided by major credit reporting agencies (credit bureaus), Equifax, Experian, and TransUnion in the U.S. Banks usually inquire at all major credit bureaus.

[6]Mortgage brokers are active in this market and account for approximately one-third of originations. Brokers have a fiduciary responsibility to act on behalf of their client and thus should have similar incentives and behavior to consumers.

Next, the lender verifies the borrower's eligibility for loan terms in an underwriting stage. This involves verifying that the information provided by the borrower is accurate by, for example, looking at tax returns or W-2 forms, or asking the borrower to explain items from their bank statement. For a mortgage, the lender additionally initiates an appraisal of the property, which is critical in determining the loan-to-value ratio. Lenders may interpret the information uncovered by this review differently and therefore differ in their assessment of applicant creditworthiness. For example, when verifying the W-2, the borrower's manager cannot be reached because they are busy. After gauging creditworthiness of the applicant, the lender can either approve a loan or reject the application. If the application is rejected, the borrower must search for another lender. To allow for shopping, credit bureaus do not disclose borrowers' recent credit applications to prospective lenders.[7] Thus lenders do not observe borrowers' application history during the shopping period. If the application is approved, the final contract terms offered to the borrower are settled at this point. The last step involves "closing" the deal where various contractual documents are signed.

Our model is an extension of the standard sequential search model proposed by Carlson and McAfee (1983) and McCall (1970) to match this institutional setting.[8] Lenders post interest rates for loans and borrowers search for these loans sequentially, incurring a constant search cost for each sampled rate. Lenders perform an in-depth credit check to obtain imperfect but informative information on applicants' creditworthiness. The lender can either approve a loan or reject the application. If an application is rejected, the borrower must search for another lender, hoping that the second lender draws a different signal of creditworthiness.

2.2. *Borrowers*

Borrowers 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: $x_h > x_l$.[9] Creditworthiness and search costs are $i.i.d$ across consumers and types.[10] Borrowers' utility from a loan at rate r is:

Similarly, our understanding is that brokers do not realize a customized rate for the borrower without each prospective lender independently pulling the borrower's credit, exactly as if the borrower were applying independently.

[7]See e.g. this article from the CFPB for more details on the shopping window. [Accessed June 5, 2023]

[8]We consider a setting with simultaneous search as in Stigler (1961) in Appendix E.

[9]The model can be extended to include continuous creditworthiness type distribution. The data only contain a binary ex post credit outcome, repayment or default, which limits our ability to identify and estimate a continuous distribution.

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

$$u(r,z) = -r + \sigma x_z.$$

Borrowers prefer loans with lower interest rates. To illustrate that standard adverse/advantageous selection does not drive our results, we allow consumers with different creditworthiness to have different preferences over obtaining a loan. Less creditworthy borrowers are more willing to take up loans if $\sigma < 0$, similar to standard adverse selection models. Conversely, if $\sigma > 0$ then more creditworthy borrowers are more willing to take up a loan, a feature generally attributed to advantageous selection models. As we will soon see, this parameter has no bearing on consumer search. We assume that interest rates do not directly affect default.[11] Realized default does not enter 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.

2.2.1. *Search*

Let $H(\tilde{r})$ be the (possibly discrete) perceived distribution of rates offered in the market with support $[\underline{r}, \bar{r}]$. Borrowers 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, they 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 loan or reject the offer and continue searching next period. Applicants are approved and end their search with probability p_z, which they take as given. They search again if their application is rejected or they choose not to apply for the loan.[12]

To characterize optimal search behavior, consider a borrower of type z with search cost c_i who was offered a loan at rate r. They will keep searching so long as their search cost c_i is smaller than the expected gain of searching once more:

$$c_i \leq \int_{\underline{r}}^{r} \underbrace{p_z}_{pr.\,approval} \underbrace{((-\tilde{r} + \sigma x_z) - (-r + \sigma x_z))}_{better\,mortgage} dH(\tilde{r})$$

[11] Including moral hazard would not change the qualitative predictions of the model, but would require additional variation in the data to estimate moral hazard.

[12] We define search to occur any time a borrower pays their search cost to sample a new rate, for which they may not be approved. This is in contrast to standard models in which a search implies that a price at which a transaction could take place has been identified.

The expected gain has two components. The first is the potential gain from finding a lower rate loan $(r - \tilde{r})$. The second is the probability the borrower will be approved for the loan p_z. This condition reduces to the standard search problem of Carlson and McAfee (1983) if borrowers are always approved ($p_z = 1$). Denote by r^*_{iz} the rate at which a borrower with search cost c_i and repayment type z would be indifferent between searching further and accepting the loan:

$$c_i = p_z \int_{\underline{r}}^{r^*_{iz}} (r^*_{iz} - \tilde{r})\, dH(\tilde{r}) \tag{1}$$

The borrower will optimally apply for any loan offered with interest rate less than or equal to r^*_{iz}, and will reject any loan offer above r^*_{iz}. As is standard in models of sequential search, reservation rates are an increasing function of search costs.[13] From the perspective of an individual borrower, the approval process exacerbates search costs. We can see this more formally by re-writing eq. (1):

$$\frac{c_i}{p_z} = \int_{\underline{r}}^{r^*_{iz}} (r^*_{iz} - r)\, dH(r) \tag{2}$$

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 $\frac{c_i}{p_z}$. The fact that they may be rejected for a loan in the future increases borrowers' willingness to accept a more expensive loan.

2.3. *Interest rate setting and loan approval (Supply)*

In the mortgage setting to which we apply the model, most rates (97.4% of our data) are offered in discrete 1/8pp increments. We therefore model the rate setting problem of lenders within a discrete choice framework. We assume risk-neutral lenders post interest rates by choosing from a menu of K discrete potential rates to offer, $r_k \in \{r_1, \ldots, r_K\}$. A fraction $\tilde{x}_z$ of a loan to type z borrowers is expected to be repaid, where $\tilde{x}_h \geq \tilde{x}_l$ by assumption. Lenders face a common expected cost m, which comprises the cost of capital as well as common regulatory and administrative costs.

We depart from the standard sequential search model by assuming that borrowers observe their creditworthiness x_z but the lender does not. Before obtaining a loan, the lender carries out an in-depth check of applicants' creditworthiness, which generates an informative but imperfect signal

[13]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.

$s_i \in \{s_h, s_l\}$. The probability that borrower is of repayment ability x_z is revealed as a high type is $p_z = Pr\{s_h|x_z\}$. The in-depth review is informative so high repayment ability borrowers are more likely to be revealed as such: $p_h \geq p_l$. We assume that applications generating signal s_h (indicating the borrower is high type) are approved, while those generating s_l are rejected.[14]

Because borrowers sort, setting the interest rate affects both the expected quantity of loans the lender underwrites and the probability of repayment on their loans. Normalizing the size of the market to 1, lenders' expected profit from charging an interest rate r is

$$\mathbb{E}[\Pi(r)] = \lambda q_h(r)(r \cdot \tilde{x}_h - m) + (1-\lambda) q_l(r)(r \cdot \tilde{x}_l - m) \tag{3}$$

where $q_z(r)$ represents the market share of type z individuals that a lender offering rate r captures.[15] Each lender j faces an additional idiosyncratic profit shock to charging specific rates $\xi_{j,k}$, which are i.i.d. and distributed according to a Type 1 Extreme Value (T1EV) distribution with scale factor σ_ξ. These $\xi_{j,k}$ represent idiosyncratic lender-rate specific shocks, such as random administrative costs, the preferences of loan officers or differences in regulatory environments. Lender j posts an interest rate to maximize its profits:

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

Since $\xi_{j,k}$ is i.i.d. T1EV, the probability that rate r_k maximizes the lender's profit is:

$$Pr\{j \text{ choose } r_k|m, \sigma_\xi\} = \frac{\exp\left(\mathbb{E}[\Pi(r_k)]/\sigma_\xi\right)}{\sum_{\tilde{k}=1}^{K} \exp\left(\mathbb{E}[\Pi(r_{\tilde{k}})]/\sigma_\xi\right)} \tag{4}$$

Specifying rate setting as a discrete problem resembling a game of mixed strategies in this way generates dispersion in posted rates so long as σ_ξ is non-zero. This approach does not change the qualitative implications of the model but guarantees equilibrium existence with adverse selection, allowing us to compute counterfactuals across a wide range of policy proposals.[16] We assume that screening is valuable, which is consistent with observing rejected applications in the loan market.

---

[14] Many credit products, including mortgages, operate using rate sheets which fix a price given a borrower's creditworthiness. Lenders have no incentives to reprice offered rates, so long as the expected repayment amount of the low creditworthiness borrower is low enough that lending to the low type is unprofitable at any legal interest rate.

[15] The profit function is specified in terms of percentage points of interest. In empirical applications, observed interest rates are frequently residualized against borrower characteristics, so that the interest rate r may take on positive or negative values. $\Pi(r)$ thus reflects 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.

[16] See, for example, Handel et al. (2015), who introduce the Riley (1979) equilibrium concept to ensure existence.

Market Shares: To construct the market share of type z individuals at interest rate $q_z(r)$, consider the probability that a type z borrower with reservation rate r^* borrows at a rate which is no higher than r. The borrower never applies for a rate above their reservation rate; thus this probability is 1 if $r^* \leq r$. If $r < r^*$, this probability is equal to the probability that the borrower is offered a rate less than or equal to r, given that they were offered a rate less than r^*. Thus,

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

One may integrate over the distribution of reservation rates to calculate the share of the type z market accounted for by lenders charging a rate less than r

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

where $F_z(r^*)$ and $f_z(r^*)$ are the distribution and density of reservation interest rates for borrowers of type z. 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 $h(r)$:

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

Intuitively, a lender charging a rate r obtains a fraction $1/H(r^*)$ of the market for borrowers with reservation rate r^*, but may only lend to individuals with reservation rates above r. This market share equation is not degenerate: price dispersion survives in equilibrium. 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 lenders (Salz, 2017):

$$\frac{dq_z(r)}{dr} = -\frac{f_z(r)}{H(r)} < 0. \tag{6}$$

Intuition: To gain intuition for lenders' decision, consider the impact of a unilateral small increase in the offered rate r on expected profits, ignoring that the rate space is in fact discrete. First, define $q(r)$ to be the share of the total market that a lender would earn by charging rate r and

$\tilde{\chi}(r)$ to be the expected recovery rate on loans originated at rate r,[17] and rearrange equation (3) to

$$\mathbb{E}[\Pi(r)] = q(r)\left[r\tilde{\chi}(r) - m\right] \tag{7}$$

The derivative of the expected profit function is therefore:

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

The marginal benefit of raising the rate is a higher profit on loans to existing borrowers. The marginal cost of raising prices has two components. First, the lender loses some market share $\frac{\partial q(r)}{\partial r} \leq 0$, because marginal borrowers now choose to keep searching instead of accepting the loan. The profits lost on each borrower are $(r\tilde{\chi}(r) - m)$. The second cost of increasing rates is that doing so may attracts weakly worse pool of borrowers: we will soon show $\frac{d\tilde{\chi}(r)}{dr} \leq 0$. The borrower pool for firms with high rates is worse because more creditworthy borrowers have lower reservation rates and are therefore less likely to apply for a loan when the price increases. This last component changes lenders' pricing incentives relative to a standard search model. In the benchmark model, search behavior and reservation rates are independent of borrowers' creditworthiness, which implies that $\frac{d\tilde{\chi}(r)}{dr} = 0$. Therefore, approvals change the lenders' pricing problem by introducing adverse selection, which decreases incentives to raise rates.

Strategic Complementarities: The presence of search and screening generates strategic complementarity in rate setting. Suppose that all lenders increase their offered rate exogenously. The first effect of such a shift is to give lenders a higher share of the market at every potential offered rate r_k as borrowers select away from high priced lenders ($H(r)$ appears in the denominator of the market share equation (5)). As a result, lenders can charge higher interest rates without sacrificing market share. Second, the outward shift in $H(r)$ changes the mix of borrowers that sort to a lender charging r_k. Because low type borrowers have higher reservation rates than high type borrowers, an outward shift in $H(r)$ weakly increases the high type share of borrowers at each rate r_k. This introduces a novel force for strategic complementarity into our model: as the offered distribution of rates shifts rightward, lenders' expected recovery rates increases at every rate.

[17] The total market share is the average of market shares of the two types: $q(r) = \lambda q_h(r) + (1-\lambda)q_l(r)$. The expected recovery rate is likewise the average of type-specific recovery rates weighted by the share of borrowers who are each type at rate r: $\tilde{\chi}(r) = \tilde{x}_h\tilde{\lambda}(r) + \tilde{x}_h(1-\tilde{\lambda}(r))$, for $\tilde{\lambda}(r) = \lambda q_h(r)/q(r)$ the share of borrowers at rate r who are high type.

The market distribution of rates influences type-specific reservation rate distributions, as implied by equation (1). As the distribution of rates shifts rightward, borrowers become more willing to apply for high interest rates, which feeds into lenders' residual demand curves from the two types $q_z(r)$. Search frictions and screening therefore generates an amplifying strategic complementarity: shifts in the distribution of rates (e.g. due to shifts in the cost of funds m) lead to shifts in the distribution of reservation rates for the two types, which feed back into lenders' optimal rate setting. Likewise, shifts in the distribution of reservation rates (e.g. due to changes in screening technology) affect the distribution of offered rates which again shifts borrowers' reservation rates. Section 6 shows that this amplification is quantitatively important.

2.4. *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 $\{G(c), \lambda, p_h, p_l, x_h, x_l, \sigma_\xi, m\}$,

1. $H(r)$ is consistent with the lender profit maximization of problem (4).
2. The reservation rate strategies satisfy equation (1).
3. Market shares of high and low types, $q_h(r)$ and $q_l(r)$, are calculated according to equation (5) and integrate to one; i.e.
$$\int q_z(r)dH(r) = 1 \qquad z \in l, h$$

A description of our approach to computing equilibria is provided in Appendix section D.3.

2.5. *Model Implications*

2.5.1. *Approval Process Induced Adverse Selection*

Adverse selection arises in equilibrium through the approval process. In standard models, adverse (advantageous) selection arises because creditworthiness is negatively (positively) correlated with preference for credit. This correlation is represented by σ in our model. In our context, borrowers' search behavior is independent of this correlation as σx_z drops out of the borrower's decision.[18] Instead, the informative approval process leads to adverse selection through its

[18]Note that all borrowers will continue to search until a loan is originated due to our implicit assumption that borrowers find loans worthwhile. If borrowers instead had some outside option to not receiving a loan, different values of σ may correlate with different realized shares of high and low types in the population: σ may affect the equilibrium value of λ or total market size. Our focus is on the search behavior of borrowers taking as given the composition of types in the market, and so we abstract from this consideration.

influence on reservation rates. Formally, consider two borrowers with the same search costs, but different creditworthiness. From equation (1), we have:

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

so that $p_h > p_l$ implies $r^*_{ih} < r^*_{il}$. That is, less creditworthy borrowers are willing to accept higher rates than more creditworthy borrowers with the same search cost. On the other hand, less creditworthy borrowers apply for expensive loans, understanding that the chances of loan approval are low in the future. Low interest rate loans attract borrowers of both high and low repayment ability. The market for expensive loans, 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: high rates are mostly taken by risky borrowers.

The necessary condition for adverse selection is that the approval process is informative, $p_l < p_h$; this replaces the standard single crossing condition in adverse/advantageous selection models. If rejection rates are the same for both types of borrowers ($p_l = p_h$), there is no adverse selection despite the underlying asymmetric information. A corollary of this observation is that changes in the screening technology which widen the gap between p_h and p_l increase the degree of equilibrium adverse selection by steepening the relationship between transacted prices and default rates.

2.5.2. *Non-Identification of search cost from transaction prices alone*

Search costs are are often estimated from transaction prices because search data are not available (e.g. Hortaçsu and Syverson (2004)). This common approach implicitly assumes that consumers can obtain the good at the listed price ($p_z = 1$). Under this assumption, the distribution of search costs is identified from transaction prices and market shares. As we have shown, search costs cannot be identified from transaction prices alone when the approval probability is less than one because approval probabilities and search costs move transaction prices isomorphically (equation (2)): a high search cost customer with a high approval probability sets the same reservation price as a low search cost customer with a low approval probability. Because the econometrician does not observe market shares for high and low types separately, one is unable to recover the distribution of search costs and approval probabilities from market shares and realized rates. Indeed, applying the standard estimation approach overestimates search costs by an order of $\frac{1}{p_z}$. This problem is therefore exacerbated for groups with low approval rates. For this reason, one may infer those with high realized interest rates to be "unsophisticated" borrowers with high search cost, when in fact they are rationally responding to low application approval. In Section 5.1 we show how the model may be estimated using data on originated loans, loan applications, loan performance and search. We show there that estimating the model using data on search but ignoring application rejections underestimates search costs, because doing so attributes all frequent-searchers as those with low search costs and ignores the possibility that frequent-searchers are routinely rejected.

2.5.3. *The relationship between search and prices*

Standard search models generate a decreasing relationship between search and transaction prices. The introduction of informative approvals can generate a non-monotonic relationship between search and transacted prices. The possibility of application rejection creates two reasons for a borrowers to continue to search. First, as usual, a borrower might draw a loan with an interest rate above their reservation rate, $r > r^*_{iz}$, and so choose not to apply for the loan. Alternatively, the borrower might discover a loan with $r \leq r^*_{iz}$ only to have their application declined. The total probability that a borrower searches again is thus:

$$Pr\left\{Search\, again\right\} = \underbrace{1 - Pr\left\{r < r^*_{iz}\right\}}_{\text{Do Not Apply}} + \underbrace{Pr\left\{r < r^*_{iz}\right\}}_{\text{Apply}}\underbrace{\left(1 - p_z\right)}_{\text{Rejected}} = 1 - H\left(r^*_{iz}\right) p_z.$$

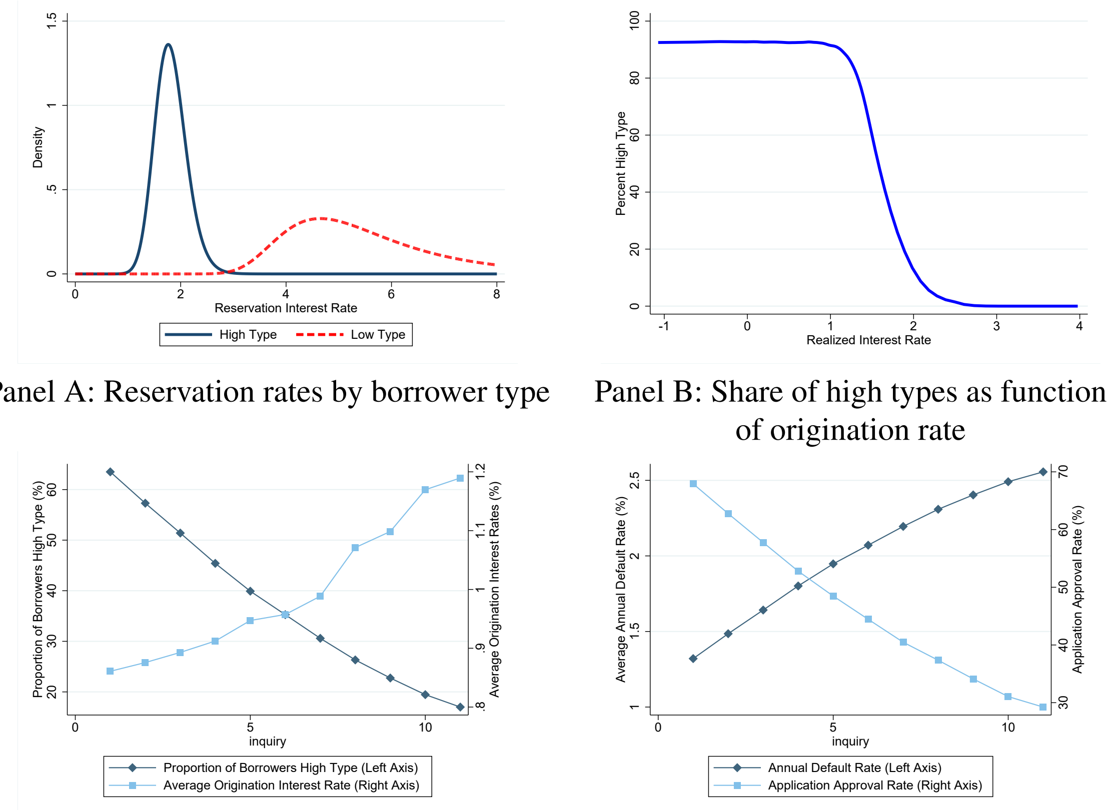
Therefore, the probability that a borrower with a reservation rate r^* searches more than s times is:

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

In the benchmark model in which there are no rejections ($p_z = 1$) or rejections are uninformative ($p_h = p_l$), transaction prices and search are negatively correlated. Low search cost (financially savvy) customers, have lower reservation rates 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.

The probability of rejection introduces two offsetting forces. Less creditworthy borrowers are more willing to accept higher rates – $H(r^*_{iz})$ is higher – which pushes them to search less. However, less creditworthy borrowers are also more likely have their applications rejected, 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 in which $p_h = 0.95$ and $p_l = 0.05$, and plot the results in Figure 1. 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

FIGURE 1.—Characteristics of a Sequential Search Model with Informative Screening



Panel A: Reservation rates by borrower type

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

Panel C: Share high type and mean rates by search

Panel D: Relationship between search and default rates/approval probabilities

Notes: Figure plots key aspects features of the model with informative screening. Data are simulated from a model in which application approval parameters are $p_h = 0.95$ 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. Panel C shows the percentage of successful borrowers who are high type as a function of search and the relationship between search and originated interes rates. Panel D displays the relationship between search and default rates and application approval probability.

search, while high type individuals dropout of the sample as they find acceptable loans. As a result, borrowers' average reservation rate increases with the number of searches. Indeed, Panel C shows a positive relationship between search and realized interest rates for this simulated sample. Borrower sorting in a search model with informative applications can therefore generate a seemingly puzzling fact that borrowers, who search more, pay higher rates on average even if lenders do not observe borrowers' past searches.[19] Rejections alone are not sufficient to explain this fact. If all borrowers

[19]Intuitively, if lenders could observe prior search they would have a direct incentive to charge higher rates to customers who search more strengthening the model prediction, but the model would lose much of its tractability.

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.

2.5.4. *Default and Approvals*

Our model suggests that search behavior should predict default behavior of borrowers ex post, conditional on observables. Defining $\hat{\lambda}(s)$ to be the share of high type borrowers among loans realized after s inquiries (loan applications), the model implies that the average default rate of borrowers with s inquiries should be $\hat{\lambda}(s)(1-x_h)+(1-\hat{\lambda}(s))(1-x_l)$. If $\hat{\lambda}(s)$ is declining in s (as in Figure 1), then borrowers with a large number of inquiries should be less likely to repay the lender ex post if $x_h > x_l$. Figure 1D illustrates this for our simulated set of borrowers in our scenario with highly informative screening.

Similarly, the probability that a loan application is approved for a borrower with s searches is $\hat{\lambda}(s)p_h + (1-\hat{\lambda}(s))p_l$. Since the type of a borrower who applies for a loan 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 the number of searches is unobservable. Figure 1D shows this decreasing relationship between application approval probability and inquiry counts for our simulated data. Note that in the baseline model, in which approvals are not informative, the default and approval probabilities are independent of the number of inquiries.

2.5.5. *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 possibly non-monotone or non-decreasing relationship between search and default
5. A possibly non-monotone or non-decreasing relationship between search and rejection
6. Placebo: Groups that are highly unlikely to have their application rejected (as p_z approaches 1) show a monotonically decreasing relationship between search and realized interest rates.

Predictions 1 and 2 are common to search models. Predictions 3-5 distinguish the model with informative screening from a benchmark model without approvals. Prediction 6 offers an additional test: if there is a segment of the population whose applications are never rejected, then this population should behave according to the benchmark search model. We verify the model's predictions using data from the mortgage market below.[20]

3. DATA AND MEASURING SEARCH USING CREDIT REPORTS

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 approximately 1.3 million mortgages originated between 2001 and 2011. At origination, we observe the borrower's credit score, the LTV ratio, the loan characteristics (origination balance and term), interest rate (inclusive of fees and points), the backend DTI ratio, whether the loan was originated through a broker, loan purpose, occupancy, and the location of the mortgaged property (zip code, MSA and state). In addition, we also have information on some of borrower's demographics, including years of school, age, gender and their monthly income at origination. Once the loan is originated, a servicer reports monthly performance until the end of our performance period, January 2015, 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.

Our second dataset contains millions of applications for mortgages intended to purchase or refinance a single family property from 2001 to 2013.[21] We term this dataset our "Application Data." 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. 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. In all our analysis, we drop all applications with more than 11 inquiries – the maximum inquiry count in the

[20]Settings outside of the mortgage market also feature search and screening. In the labor market, for instance, unemployed workers are less likely to find a job the longer they have been searching. Through the lens of our model, this arises if some workers are low productivity and employers interview to reveal workers' type. Low-type workers would both be willing to accept lower wages and have long unemployment durations in response to a low probability of converting an interview into a job. Jarosch and Pilossoph (2019) develop a model in which the long-term unemployed are statistically discriminated against because they are more likely to be low type.

[21]The shorter time period relative to application sample above reflects data sharing constraints with the GSE. We are unable to match an originated loan to its application.

loan data – on their credit report.[22] This is to be consistent with the model, in which all borrowers eventually get a loan. In addition, relative to the loan data, there may be some measurement error in the application data's inquiry counts owing to a miscoding of credit pulls among a subset of lenders. As robustness, we explicitly allow for measurement error in our structural estimation. Our final sample contains 3.26 million mortgage applications.

Table I reports summary statistics for our sample. Our data consist of prime borrowers. Therefore the average 726 FICO score of approved borrowers substantially exceeds that of the US population, which was 688 in April 2011.[23] The average combined loan-to-value (CLTV) ratio was 73.8% and average back-end debt-to-income ratio was 37.6. Applicants look observably fairly similar to realized loans, with average FICO of 728, and average CLTV of 73.3%. 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 64.6 in the application dataset. We see similarly large standard deviations in both CLTV and DTI ratios. Indeed, these loans are not without credit risk: the annualized default rate is 2.2% in our sample.[24]

3.1. *Measuring Search: Credit Application Process and Inquiries*

We measure the intensity of borrowers' mortgage search using formal credit inquiries: "total inquiries" that borrowers register with credit bureaus when applying for credit directly or through a broker. To obtain these inquiry measures, we merge our loan and application data with applicants' credit reports provided by a consumer credit bureau using applicants' social security numbers.[25] We limit the search window to within 45 days of the final mortgage application, following the credit bureau definition of search.[26] Credit bureaus entitle borrowers to a "shopping window" of 45 days. During this window, consumers' credit scores are not penalized for additional searches: multiple credit checks count as a single inquiry when computing a borrower's credit score, but each check appears in our data. Focusing on this formal search is important in our context.

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

[23] This article from FICO reports this statistics. [Retrieved Nov, 2016].

[24] Our dataset includes loans originated through the housing boom, bust and recovery. Appendix Table A1 reports summary statistics for our two datasets across three origination periods.

[25] Inquiries have been used as a measure of search subsequent to our paper, most notably Ambokar and Samaee (2019b).

[26] If a borrower performs a credit check with multiple credit bureaus, that counts as a single inquiry.

TABLE I

SUMMARY STATISTICS FOR MORTGAGES AND APPLICATIONS

	Loan Data		**Application Data**	
	Mean	SD	Mean	SD
# Inquiries	2.61	2.00	5.47	3.05
$Pr\{$Approval$\}$ (%)	–	–	86.70	–
Origination Interest Rate (%)	5.69	0.86	–	–
FICO	725.8	62.5	727.9	64.6
CLTV	73.8	18.4	73.3	19.1
Back-end DTI ratio	37.6	12.8	35.8	12.8
$Pr\{$Default$\}$ (Annualized %)	2.21	–	–	–
$Pr\{$90+ Days Delinquent$\}$ (Annualized %)	1.34	–	–	–
Observations	1,316,807		3,263,680	

Notes: The first two columns report 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.

Alternative sources of data – especially survey based data – may not adequately capture search that ends in application rejection. Indeed, in the oft-cited NSMO data, a number of borrowers report "seriously considering" fewer lenders than they applied to.[27] This suggests that respondents may not count rejected applications as "serious considerations" ex post. The average realized loan has 2.6 inquiries at the time of origination, while the average application has 5.5 inquiries, reflecting the fact that not all applications are originated.

Total inquiries might overstate mortgage search if borrowers search for other credit products during the same window. We check whether non-mortgage inquiries contaminate total inquiries in two ways. First, we measure the share of mortgage-related (as determined by the credit bureau) inquiries as a proportion of total inquiries for a given borrower in the month prior to their mortgage origination, for which we have inquiry purpose data. More than 80% of total inquiries during this period are mortgage related. Given it usually takes more than one month from the original inquiry to close a mortgage, the true share is likely to be higher. Importantly, the distribution of mortgage-

[27]The survey's main questions surrounding search are 1) "How many different lenders/mortgage brokers did you seriously consider before choosing where to apply for this mortgage?" and 2) "How many brokers/lenders did you end up applying to?" Appendix B.2 shows that more than half of borrowers who applied to 5+ brokers/lenders report that they only considered four or fewer brokers/lenders. Borrowers who report applying to 3, 4 or 5+ lenders report considering fewer lenders than they applied to on average. Furthermore, respondents likely do not "seriously consider" rates they find through soft search activities such as internet searches.

related inquiries in the raw credit bureau data matches the distribution of total inquiries in our loan-level data, suggesting that most inquiries around this time are mortgage related.

Second, we find no credit limit increases for credit cards or home equity lines of credit (HELOCs), on average, in both the month of mortgage origination nor in the month preceding origination. This suggests that consumers' search for non-mortgage credit is limited during the period over which we examine inquiries. Since credit scores are adversely impacted when borrowers take up credit products, there are strong incentives not to formally search for other credit products before applying for a mortgage.

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. This "soft search" does not guarantee mortgage origination at a specific rate: 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 pulling the borrower's credit report ("an inquiry") and knowing the house characteristics. Consequently, our measure captures borrower search over formal terms, which is what we model. One way to think of soft search is that borrowers are learning about the distribution of prices, rather than specific prices offered by lenders (Rothschild, 1974). Since empirical applications typically do not observe soft search, including in survey data on serious rate considerations, we omit this feature from the model and empirical work.

In Appendix B, we benchmark the inquiry measure against several other datasets to validate our assertion that inquiries are a good proxy for search in the mortgage market. First, we show that the distribution of inquiries in our data matches that from credit report data. Second, we use HMDA data to show that this distribution of inquiries can be generated in a simple back-of-the-envelope calculation in which search only occurs if borrowers' mortgage applications are rejected or they decide not to take the mortgage up. Third, we study the National Survey of Mortgage Originations (NSMO) and find that there is ambiguity in the appropriate measure of search in the survey data.

4. QUALITATIVE MODEL PREDICTIONS.

In this Section, we test the qualitative predictions of our model outlined in Section 2.5.

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 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 in our loan data across three origination time periods (Figure A1A), and three different FICO based creditworthiness subsets (Figure A1B). There is substantial mortgage rate dispersion within every subset, with interest rates differing over 3 percentage points (pp) within each group. We purge observable differences between borrowers using the following specification:

$$r_{itm} = X_i'\beta + \mu_t + \mu_m + \varepsilon_{itm}, \tag{8}$$

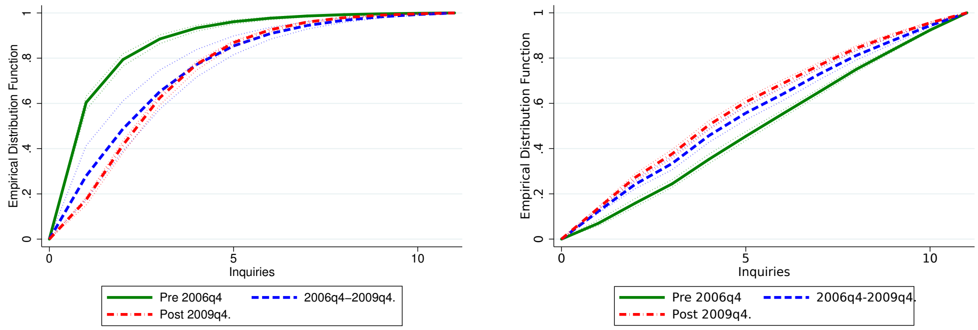
in which r_{itm} represents the origination rate of borrower i at time t in market m. X_i are borrower and loan characteristics; specifically the FICO score, LTV, DTI, income, years of education, the type of the mortgage (term, ARM vs FRM, purchase or refinance), and whether the borrower is an investor. We compare borrowers in the same market at the same point in time by including state fixed effects μ_m and time fixed effects μ_t. Our data set was collected by the lender for the purposes of making the loan and selling it to GSEs. Thus, the controls we observe and use closely approximate the variables used to set rates: the R^2 from the above regression is 0.796. Figure A1C plots the distribution of the regression residuals.

A substantial amount of residual rate dispersion remains, with a $90^{th} - 10^{th}$ percentile difference of $0.9pp$. At the average loan amount of \$169 thousand, this difference results in \$1,080 larger mortgage cost per year. Our estimate of the standard deviation of residual price dispersion of 39bp are in the range of 27bp found in Ambokar and Samaee (2019b), 50bp in Allen et al. (2014), 50bp in Alexandrov and Koulayev (2017), and 54bp in Bhutta et al. (2021). 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. As predicted by our model and established by the existing literature, observably similar borrowers pay substantially different mortgage rates.

4.2. *Borrower Search, Sophistication, and Creditworthiness*

Price dispersion has been documented in several other mortgage markets, but there is little direct measurement of search behavior. To highlight that borrower creditworthiness plays a central role in observed search behavior, we plot the search distribution across FICO levels in our loan (Figure 2A) and application data (Figure 2B). The median borrower who obtains a mortgage does not search much, having only 2 inquiries on her record (Figure 2, Panel A). A borrower at the 75^{th} percentile searches 3 times. Mortgage applicants search substantially more, with a median of 5 (Panel B). This result suggests that borrowers who frequently search are less likely to be approved.

FIGURE 2.—Inquiry distribution among mortgage borrowers and applicants



Panel A: Loan-Level Dataset Panel B: Applicant Dataset

Notes: Figure plots distribution of inquiries in our loan (Panel A) and application (Panel B) datasets 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. Note we drop those with over 11 inquiries in the application data.

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

$$s_{itm} = X_i'\beta + \mu_{mt} + \varepsilon_{itm} \tag{9}$$

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

We present the results in Appendix Table A2. Importantly, more creditworthy borrowers search less conditional on other characteristics. 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. If FICO proxied only for financial sophistication, one would expect the opposite: low FICO borrowers should search less, not more.[28] Borrower characteristics such as education and race are correlated with search, but the simple correlations are inconsistent with the intuition that sophisticated borrowers search more. College-educated borrowers, traditionally considered sophisticated (Woodward and Hall,

[28]The FICO score is a measure of creditworthiness, but has also been used as a measure of consumer sophistication.

2012; Gurun et al., 2016), have 0.11 fewer inquiries than non-college borrowers at the time of mortgage origination.

This distribution of inquiries aligns closely with what one would expect given the frequency of application rejection in other administrative data. The Home Mortgage Disclosure Act (HMDA) requires lenders to report all applications for mortgage financing, along with their outcome. In the HMDA data, 20-30% of applications are rejected and a further 15% of approved applications are not taken up by the borrower, suggesting a 60% origination probability. Assuming originations are i.i.d., a 60% origination probability for applications would imply that 60% of originated mortgages have 1 inquiry on record, 24% (0.4*0.6) have 2 inquiries, 9.6% have 3 inquiries, 3.8% have 4 inquiries, leaving 2.6% to have 5+ inquiries. This simple back-of-the-envelope exercise likely under-represents the right tail of the inquiry distribution for two reasons: borrowers can elect not to apply for a loan after receiving a quote (for which an inquiry is necessary) and the probability of originating a loan is not i.i.d. because who have been rejected once are more likely to be rejected again.[29] Nevertheless, this distribution is fairly similar to the distribution of inquiries in our loan dataset, suggesting that our inquiry data captures an important element of search – continued search after application rejection – that may not be captured by surveys.

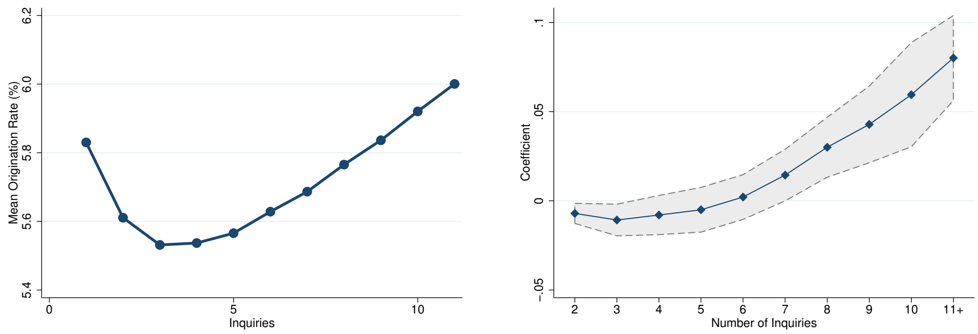
4.3. *Do Borrowers Who Search More Obtain Cheaper Mortgages?*

This section presents a new robust fact: origination rates do not decline monotonically with search. Panel A of Figure 3 plots the average mortgage rate as a function of search for borrowers. In the canonical search model, the average price (rate) monotonically declines with search. As the number of searches increases from one to three, the interest rate indeed declines. However, after three inquiries, additional search is correlated with increased mortgage rates. Frequent-searchers obtain more expensive mortgages than borrowers in the middle of the search distribution. The same pattern persists among those with low, middle and high FICO scores (Appendix Figure A2).

This pattern continues to hold within a regression framework controlling for differences across markets and borrower characteristics. We estimate the following regression

[29]Indeed, if we limit attention to a subsample of risky borrowers - those with DTI above 50% and CLTV above 95% - we find that only 48% of applications are originated. The same back-of-the-envelope would imply that, among this group, 48% have 1 inquiry, 25% (0.48*0.52) have 2 inquiries, 13% have 3, 7% have 4, and 7% have 5+. Note further that nearly all of these observably risky borrowers take up an approved mortgage application, which provides reduced form evidence for the model's core mechanism: low type borrowers are, in some sense, "less picky." We detail this analysis and suggest reasons why survey data may understate search after rejection in Appendix B.

FIGURE 3.—Relationship Between Search and Mortgage Origination Rates



Panel A: Raw | Panel B: Conditional on Covariates

Notes: Figure plots relationship between origination interest rates and search in our loan data. Panel A plots the raw relationship between search and origination rates, while Panel B plots regression coefficients estimated from equation (10) using OLS across three FICO sub-samples. The dependent variable in each regression is the origination interest rate plus points and fees on a loan. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. Standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.

$$r_{itm} = \sum_{s \geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_t + \mu_m + X_i' \gamma + \varepsilon_{itm} \tag{10}$$

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 inclusive of fees. The independent variable of interest is the amount of search the borrower undertook before taking up a mortgage s_i. The coefficients of interest β_s measure the mean change in mortgage rates for a borrower who searched s times, relative to a borrower who only searched once. To ensure that the correlation between search and mortgage rates is not driven by borrower or mortgage characteristics, we include the same extensive controls in X_i as in equation (8), such as the borrower's FICO score, LTV, etc. By including time fixed effect μ_t, we absorb any aggregate fluctuations, such as changes in the risk premia, while the location fixed effect μ_m absorb persistent differences in local supply and demand across markets. We cluster standard errors at the state $\times$ origination quarter level. In effect, we consider two borrowers in the same location, at the same point in time, with the same observable characteristics, and compare how the interest rate charged on their mortgage differs with the amount of search.

Panel B of Figure 3 plots the coefficients β_s. As the figure suggests, borrower, location, or time differences do not drive the relationship between search and interest rates. Increased search has a U-shaped, or even monotonically increasing relationship with interest rates. These results persist if we estimate equation (10) within populations defined by observable borrower characteristics, as shown in Appendix Figures A2 and A3. They hold controlling for a richer set of covariates – namely the set of loan-level price adjustment (LLPA) factors used by Fannie Mae – for low-, middle- and high-education populations, for black, white and Hispanic borrowers, for low-, middle- and high-income borrowers, as well as for refinance and purchase loans, for loans that both were and were not originated by a broker, and for loans that did/did not default ex post. In each case, there is a U-shaped or positive relationship between search and interest rates in the data.

Consistent with the model of rejection, borrowers who search a lot obtain higher rate mortgages than those who search little. We reject the prediction from standard search models that more search is correlated with lower mortgage rates.[30]

4.4. *Loan Performance and Search*

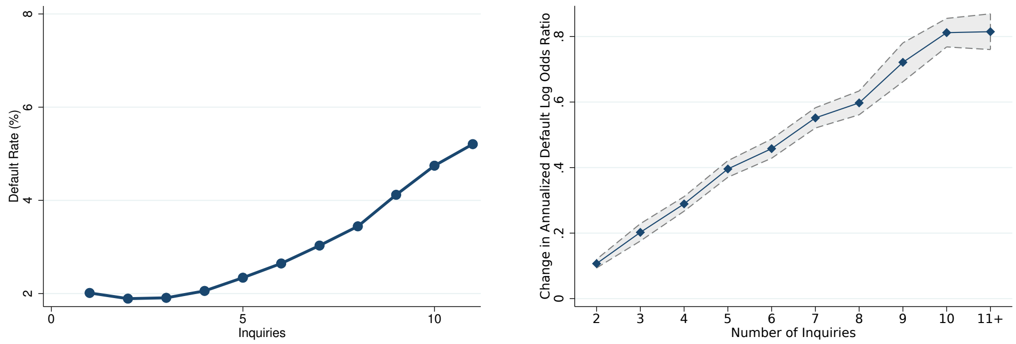
Our model predicts that less creditworthy borrowers search more in equilibrium, leading to positive relationship between search and ex-post default rates. We only observe default behavior as of January 2015 in our loan data. We assume that there is a constant proportional hazard of default for all loans. Let d_{iTm} be the probability that loan i originated T years before January 2015 in market m will default in year t having survived through year $t-1$: that is, d_{iTm} is the annualized hazard rate of default. Figure 4 confirms the positive relationship between ex post creditworthiness and search by plotting the annualized default rate (Panel A) against the number of inquiries on record for all borrowers in our sample. In the model, borrowers who search more are more likely to default even conditional information observable to the lender. We test this prediction by assuming that the annualized default rate has a logistic form, that is:

$$d_{iTm} = \frac{\exp\left(\sum_{s\geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_T + \mu_m + \gamma X_i + \varepsilon_{itm}\right)}{1 + \exp\left(\sum_{s\geq 2} \beta_s \mathbf{1}\{s_i = s\} + \mu_T + \mu_m + \gamma X_i + \varepsilon_{itm}\right)} \tag{11}$$

We estimate the parameters $(\mu_m, \mu_T, \gamma, \beta_s)$ through maximum likelihood, clustering standard errors at the origination quarter level. We define default to include both defaults and 90 day

[30] As additional evidence for the model's mechanism, we turn to data from the National Survey of Mortgage Originations (NSMO). Twenty-eight percent of borrowers report searching across multiple lenders due to concerns that they may not qualify for a loan. Those that do realize statistically significantly higher interest rates on average than do borrowers who seriously consider multiple lenders/brokers for other reasons. See Appendix B.2 for details.

FIGURE 4.—Search and Annualized Default Rate



Panel A: Raw Panel B: Conditional on covariates

Notes: Figure plots relationship between default and search in our loan data. Panel A plots the raw relationship between annualized default rates and search, while panel B plots regression coefficients estimated from equation (11) using MLE. The coefficients reflect changes in the log odds ratio of the annual default hazard relative to borrowers with one inquiry. Default is defined by the loan being at least 90 days delinquent, or entering foreclosure. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination equals 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 DTI (debt-to-income) 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.

delinquency on their mortgage payments as of January 2015. The independent variable of interest is the amount of search the borrower undertook before taking up a mortgage, s_i. The coefficients of interest β_s measure the difference in log odds ratio of default for borrowers who search s times compared with those who search just once. As with our interest rate regressions, we control for observable characteristics X_i, and include location fixed effects μ_m and time fixed effects μ_t, which absorb any aggregate fluctuations, such as changes in the regulatory environment.

Consistent with our model, frequent-searchers are more likely to default or become delinquent on their loans, even conditional on observable characteristics (Figure 4, Panel B). Our estimates imply that a borrower with 5 inquiries is approximately $e^{0.4} - 1 =$49% more likely to default on their mortgage in a given year than is a borrower with 1 inquiry, conditional on observables. This positive relationship between search and default probabilities is highly robust. We re-estimate the specification in sub-populations of low, middle and high FICO borrowers; low, middle and highly educated populations; for brokered and unbrokered loans; for refinance and purchase loans; and for low, middle, and high income borrowers (Appendix Figure A4). 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. These unobservable risks reflect anything that are not perfectly predicted by observables at the time of origination, such as income, marital or health risks which are difficult for a lender to verify.

4.5. *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. Figure C1 illustrates the strong negative correlation between search and the probability of mortgage approval.[31]

4.6. *Placebo: Borrowers who are never rejected*

Our model's predictions are consistent with the data on mortgage pricing, default, and approvals across multiple subsamples. One potential alternative explanation is that creditworthiness is observable to the lender but not the researcher, while 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: borrowers who search more will borrow at lower rates on average. Therefore, for any subset of borrowers who do not expect to be rejected, the relationship between average rates paid and search should be negative. This subsample serves as a placebo for our proposed mechanism. If, on the other hand, search is a proxy for creditworthiness observed by the lender, then we should find a non-negative relationship within this subset, as in the whole sample.

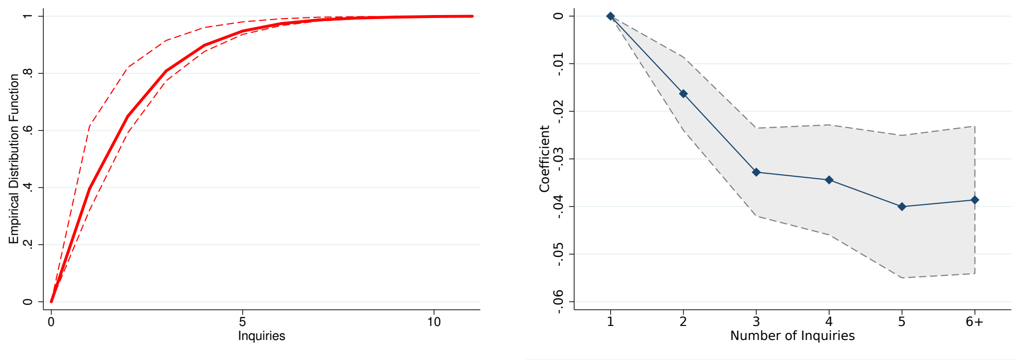
We use borrower, mortgage, location, and time characteristics to predict the probability that an application is accepted by estimating a logistic regression, and select borrowers whose mortgage applications are rejected very rarely: those with predicted approval probability greater than 97.5%. The average approval rate of this sample is 98.5%, much higher than the average approval rate of 82.2% or 89.7% for high (about 720) FICO score borrowers.[32]

---

[31] Because the application data do not contain as rich covariates, we are unable to replicate this plot conditional on observables and thus emphasize the results on interest rates and default as evidence for our model's mechanism.

[32] An earlier version of this paper showed 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%. The conditional patterns found here are also visible unconditionally.

FIGURE 5.—Search Behavior of Rarely Rejected Borrowers



Panel A: Search distribution

Panel B: Relationship between search and rates controlling for observables

Notes: Figure plots key aspects of search behavior for a pool of borrowers in our loan data 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 distribution of inquiries for these borrowers. Dashed lines plot bootstrapped 95% confidence intervals, in which bootstraps are clustered at the origination quarter level. Panel B plots regression coefficients estimated from equation (10) using OLS. The dependent variable in the regression is the origination interest rate plus points and fees on a loan. The independent variables are a set of dummy variables equal to one if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \ldots, 11+\}$. The omitted category is $s = 1$. White heteroskedasticity robust standard errors are clustered at the origination quarter level. Gray bands indicate 95% confidence intervals.

Panel A of Figure 5 shows that, despite the absence of rejections, these borrowers search: roughly 60% have multiple inquiries. These borrowers have substantial variation in realized rates: the standard deviation of residualized interest rates is 29bp for this subsample, which is slightly smaller than for the full sample.[33] However the nature of this search behavior is radically different to that found in the full sample of borrowers. We replicate the regression of equation (10) for this set of borrowers and plot the residual relationship between mortgage origination rates and search in Figure 5B. Consistent with the model, 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. In the absence of application rejection, the data support the benchmark search model, as predicted by our model when $p_h = p_l = 1$. 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.

[33]Without residualizing against observables, the standard deviation of realized rates for this sample is 71bp compared with 86bp in the full sample.

5. MODEL ESTIMATION

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 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.

5.1. *Estimation*

As we show in Section 2.5.2, traditional methods of identification in search models are insufficient in the presence of application rejections. The central challenge is that borrowers who pay high prices may either have high search costs or high probability of application rejection. However, one can estimate the model given data on search, interest rates and default. Below, we briefly outline our maximum likelihood approach, relegating details to Appendix D.

We observe the joint distribution of search S, rates R, and default D for realized loans, the distribution of search among applications, as well as observable loan and borrower characteristics X. To ensure comparability of loans in our estimation, we residualize observed rates against observable characteristics following regression equation (8). On the demand side, one must uniquely recover the set of parameters $\theta \equiv \{p_h, p_l, x_h, x_l, \lambda, H(r), G(c)\}$ given the distribution of S_i conditional on application and the joint distribution of (S_i, R_i, D_i) conditional on origination. We estimate the parameters of the lender problem $\theta^S \equiv \{m, \sigma_\xi\}$ by imposing that the offered rate distribution $H(r)$ estimated from the demand side is consistent with profit maximization.

Likelihood Construction: The probability of a loan being originated is equal to the probability that the borrower draws a rate below their reservation rate and that their application is approved. Because search is i.i.d., the probability that a loan is originated by a type z borrower with reservation rate r^* after s inquiries is

$$Pr\{\text{Originated after } s \text{ searches}|z, r^*\} = p_z H(r^*)(1 - p_z H(r^*))^{s-1} \tag{12}$$

The relationship between default and search reveals how the share of high types evolves through the search distribution. This helps identify the gap between x_h and x_l as well as p_h relative to p_l. To incorporate information on payment status as of January 2015, we assume that defaults occur with a constant hazard. Specifically, let T be the loan term in months and let t be the number of months since origination. A borrower of type z, who has seen a share t/T of his loan term elapsed by January 2015, realizes $D_i = 0$ with probability $x_z^{t/T}$ and $D_i = 1$ with probability $1 - x_z^{t/T}$. Thus the likelihood of the joint distribution of our loan data $(S_i, R_i, D_i|\theta, t, T)$ is:

$$l^{LOAN}(R_i, S_i, D_i|\theta, t, T) = \lambda \overbrace{\left(D_i(1-x_h^{t/T}) + (1-D_i)x_h^{t/T}\right)}^{\text{Probability of observing default } D_i \text{ of } h} \int_{R_i}^{\infty} \overbrace{p_h h(R_i)\left(1-p_h H(r^*)\right)^{s-1}}^{\Pr\{\text{Loan to } h \text{ with } r^* \text{ has } s \text{ inquiries}\}} dF_h(r^*)$$
$$+ (1-\lambda)\left(D_i(1-x_l^{t/T}) + (1-D_i)x_l^{t/T}\right)\int_{R_i}^{\infty} p_l h(R_i)\left(1-p_l H(r^*)\right)^{s-1} dF_l(r^*) \quad (13)$$

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

Next, we construct the likelihood of our application dataset. The probability that a borrower applies for a loan on the s^{th} inquiry is

$$Pr\{\text{Apply on } s^{th} \text{ search}|z, r^*\} = H(r^*)(1-p_z H(r^*))^{s-1} \quad (14)$$

A comparison of equation (14) with equation (12) highlights that the gap in the distribution of search among applications and originations is informative about the application approval probability p_z. The likelihood of an application conditional on being in our data is constructed by integrating equation (14) with respect to type z and reservation rates r^* and dividing by the probability that the application is in our data. Appendix D shows that this may be expressed as:

$$l^{APP}(S_i = s|\text{Applied}, \theta\} = \frac{1}{\lambda/p_h + (1-\lambda)/p_l}\left[\lambda \int H(r^*)\left(1-p_h H(r^*)\right)^{s-1} dF_h(r^*)\right.$$
$$\left. + (1-\lambda)\int H(r^*)\left(1-p_l H(r^*)\right)^{s-1} dF_l(r^*)\right] \quad (15)$$

where $\lambda/p_h + (1-\lambda)/p_l$ is the probability that a search leads to an application.

Parametric Assumptions: Although well-defined, maximizing this likelihood remains difficult. Given two joint distributions, we must estimate five parameters associated with the type distribution, default and acceptance probabilities, as well as three distributions: the offered rate distribution $H(\cdot)$, and the reservation rate distributions for high and low types, $F_h(r^*)$ and $F_l(r^*)$, respectively. To ease estimation, we assume that the offered rate distribution is well-approximated by a normally distributed random variabls parameterized by $\beta_H \equiv \{\mu_H, \sigma_H, \pi_H\}$, while the search cost distribution is well-approximated by a log-normally distributed random variable parameterized by $\beta_G \equiv \{\mu_G, \sigma_G, \pi_G\}$. These assumptions permit analytical computation of the reservation rate distribution for high and low type individuals (Appendix D.1), and is motivated by the roughly normal distribution of residualized realized rates observed in Figure A1.[34]

Differences in application and loan-level datasets: To estimate our parameters, we maximize the log likelihood for our sample of loans and applications. We do not need to observe a random sample of approved loans. We allow for selection in the sampling of approved loans based on observed characteristics. Specifically, we assume that an approved loan application is reported in our loan-level dataset with probability $q(X_i)$. We consider $q(X_i)$ to be a nuisance parameter whose estimation is not of interest. By letting this probability depend on observables X_i, we control for differences in observables between application and loan datasets. Let the set of observations in the realized loan dataset be given by $\mathscr{L}$, while the set of observations in the application dataset be given by $\mathscr{A}$. We maximize the following log-likelihood with respect to a choice of θ:

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

where $l^{LOAN}(R_i, D_i, S_i|\theta, t, T)$ is given by equation (13), and $l^{APP}(S_i = s|\text{Applied}, \theta)$ is given by equation (15). 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.

Intuition for Identification: The model is non-linear and all features of the data therefore contribute to estimating each parameter. Nevertheless, some moments in the data are intuitively more related to specific parameters. First, the difference between the distribution of search in the application and realized loan datasets identifies the level of the application approval parameters p_z. Intuitively, if applications are always approved ($p_z = 1$), then the distribution of search for applications should be identical to that for realized loans. The joint distribution of search, interest rates and default informs the share of borrowers who are high type at each level of search, which helps identify the difference in acceptance and default probabilities for the two borrower types. The steeper is the slope between search and default rates, the larger the gap in application approval and default rates between the two borrower types. Likewise, the steeper the upward relationship between search and interest rates, the larger the gap in application approval probabilities and, therefore, reservation rate distributions between the two types. Meanwhile, the higher the average

[34] We found that allowing distributions to be approximated by finite mixtures of normals or log-normals did not substantially alter our quantitative results, but led to large standard errors and most weight placed on just one of the mixtures.

realized interest rate and the more dispersed the interest rate distribution, the higher will be the estimated mean and standard deviation of the offered rate distribution. Finally, the more search that is observed, especially for those who realize low interest rates, the lower will be inferred search costs.

Supply Side Parameters: To estimate the cost of making loans m and the variance of lender profit shocks σ_ξ, we impose that the distribution of offered rates $H(r)$ estimated from the demand-side maximum likelihood routine is consistent with lenders equilibrium rate setting behavior. 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 profit-maximizing probability distribution implied by equation (4).

5.2. *Results*

Data Fit: Despite its simplicity, Figure A5 shows that the estimated model matches observed price dispersion and distribution of searches, as well as the increasing relationship between search and both default and realized (residualized) interest rates documented in section 4.

Screening Technology and Adverse Selection: Our maximum likelihood estimates are reported in Table II. Our estimates suggest that 73% of potential borrowers are low-type. Assuming a constant default hazard on a 30 year mortgage, the annualized default rate of low-type borrowers is $1 - 0.41^{1/30} = 2.9\%$. In expectation, low-type borrowers repay 66 cents of principal on a dollar. The remaining 27% are high types, who repay almost certainly. Given that lending to a bad type is extremely costly, lenders have high incentives to screen the borrowers. Our estimates suggest that lenders make few mistakes when screening high types: p_h is close to 1, so these borrowers rarely generate a bad credit signal. That is intuitive, since a bad credit check generally requires the revelation of bad information. However, the screening process is imperfect: $p_l = 0.19$ suggests that lenders do not uncover the bad information on low types in 19% of cases.

The difference between p_h and p_l of 0.807 suggest that the screening technology is very informative. In other words, lenders on average do a good job of verifying borrowers' income and employment, or conducting house price assessments and other checks which are not a part of the standard measures of borrower creditworthiness such as FICO, LTV, DTI and reported income. A simple back of the envelope suggests that the expected loss on a bad borrower applying is lowered by approximately 81% from $34pp$ (one minus the expected repayment of a low type) to $19\% * 34pp = 6.5pp$. 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.

TABLE II

MAXIMUM LIKELIHOOD ESTIMATES FOR OUR FULL SAMPLE OF LOANS AND APPLICATIONS

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

Notes: Table reports estimated model parameters obtained from maximum likelihood estimation described in section 5. Estimation employs both loan and application data. 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 informative screening technology provides a large force towards adverse selection. Low creditworthiness borrowers behave as if their search costs are $\frac{1}{19\%} = 5.3$ times those of good borrowers (eq. 2) 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 A5E. At the mean origination interest rate, the annualized default probability is 3.2% and the derivative of this default rate with respect to the interest rate paid is 1.5. Small increases in the realized interest rate lead to sizable increases in the default probability at the mean realized rate. Changing the approval process, either due to place-based government policies in credit markets or technological innovation in screening, can induce substantial changes in the extent of adverse selection. This in turn would affect the prices at which borrowers across the creditworthiness spectrum can borrow, as well as the search effort they expend in equilibrium. We examine these changes in approval policies and technology in Section 6.

Search Costs: The mean of the search cost distribution is estimated at 29.7bp.[35] One can translate these search costs into dollar terms using a mortgage calculator. 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 typical monthly payment for this loan is given by $y = Y\left(r(1+r)^T\right)/\left((1+r)^T - 1\right)$. This implies that the monthly payment on a 30-year fixed rate mortgage with principal of \$170,000 and interest rate of 4% per year – the mean mortgage in the data – is \$811.61. If a borrower with search cost c searches one additional time, they would

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

pay the equivalent of c additional basis points of interest. At the mean search cost of 29.7bp, this estimate would translate into a monthly payment increase of \$29.45. This sums to \$1,800 over five years, or \$10,603 over the life of the loan.[36] Our estimates of average costs are in line with the 27.2bp in Allen et al. (2014), and \$29 monthly in Allen et al. (2019) for the Canadian insured mortgage market. The standard deviation of 11.8bp is smaller than 23bp in Allen et al. (2014).[37]

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. (2014). To gauge whether this is sensible, we approximate the lending cost of lenders as the rate on 10-year treasury bills and compare it 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.

Bias in search cost estimation: An important result of Section 2.5.2 is that estimating the model using the standard approach of using only price and quantity data (Hortaçsu and Syverson (2004)) would overestimate search costs by an order of $\frac{1}{p_z}$. If one were to naively suppose there were one type of borrower who was rejected at the average rate in HMDA (around 25%), this would lead one to substantially overestimate search costs by factor of approximately 1/3. Approximately 50% of applications from borrowers with high DTI and LTV are rejected (see Appendix Figure B1), suggesting standard approaches would estimate search costs to be roughly twice as high as they really are for these borrowers.

Alternatively, one could re-estimate the model following the routine above to incorporate search data, but ignore rejections. To illustrate the bias that this would induce, we re-estimate our model without rejections by imposing that $p_h = p_l = \lambda = 1$. Doing so reduces the estimated mean of the

---

[36] This estimate is an upper bound of total cost in that it assumes the mortgage is never refinanced or prepaid and does not account for any discounting. Note further that very frequent searchers may be those with lower search costs than average. Borrowers with 10+ inquiries account for 1% of the loan data and the 1^{st} percentile of estimated search costs corresponds to a five-year cost of \$673.

[37] If there were more than two creditworthiness types, some of the offered rate variation across types would be explained by borrower riskiness. This could in principle reduce estimated search costs as the within-type benefit of additional searches falls with the variance of offered rates. Estimating on the never-reject sample, who are of relatively homogeneous risk profiles, yields a higher estimated offered rate variation, suggesting this concern is unlikely to be first-order.

search cost distribution to 15.4bp, around half our baseline estimate of 29.7bp. The estimated standard deviation of search costs falls to 7.5bp. Intuitively, without modeling rejection, the estimator confuses borrowers who are frequently rejected with low search cost borrowers, as both groups accumulate many searches. In sum, not accounting for rejection overestimates search cost if only price and quantity data are used, but underestimates them when search data is added.

Robustness – As detailed in the Appendix, these results are qualitatively robust to various alternative estimation choices. First, we estimate only using observations with no more than 6 inquiries, fearing high inquiry counts could be due to some sort of measurement error. Doing so leads to an estimate of average search cost of 22.5bp and screening technology $p_h - p_l$ of 0.57. It is unsurprising that arbitrarily cutting the distribution of search at a low level leads to higher estimated approval probabilities for low types; nevertheless, we find it heartening that each of these estimations yield qualitatively similar estimated search costs and a substantial role for application rejection. Second, we re-estimate the model allowing for parametric measurement error in the application data inquiry counts as described in Appendix D. The mean search cost is estimated to range between 19.3bp and 32.8bp, while the screening technology $p_h - p_l$ ranges from 0.52 to 0.60, depending on the amount of measurement error we assume.[38] Finally, we re-estimate the model on the sample of observably rarely-rejected borrowers. Consistent with intuition, we find that 99% of these borrowers are high-type, both high- and low-types repay their loans with a greater than 90% probability, the variance of offered rates is larger than in the full sample at 0.72, and average search costs are similar amongst this group as in the full sample.

Price Discrimination – We have abstracted from price discrimination in both theory and estimation. This is reasonable for two reasons. First, we have residualized realized interest rates against the set of observables that banks may use to price discriminate. Indeed, lenders often use statistical models to set prices and GSEs use rate sheets to price their loans. Second, anti-discrimination and usury laws often constrain discrimination by directly targeting prices.

Nevertheless, there may be some additional price discrimination that is unobservable to the econometrician and may bias our estimation. Intuitively, if lenders increased prices after observing a bad signal rather than rejecting borrowers, then low-type borrowers would have higher reservation rates than high-type borrowers for a given search cost, because they would

[38] We also estimate our model on a variety of subsamples and present the results in Appendix Figure A6. Estimating on subsamples suffers from power issues, as some subsamples do not constitute a large share of GSE loans. Nevertheless, the results are qualitatively intuitive: for example, observably risky borrowers have higher default rates.

effectively be drawing from a distribution of higher offered rates. To rationalize the upward sloping relationship between rates and search, a model with price discrimination would therefore have to rely on low-type borrowers having lower search costs than high-type borrowers. Furthermore, to rationalize the coexistence of substantial price dispersion and some borrowers with a large amount of search, a model with price discrimination alone would require more dispersion in search costs. Thus, by abstracting from price discrimination, it is possible that we introduce a downward bias in the variance of search costs and application approval probabilities. However, this bias is likely small in our setting for the reasons described above.

6. COUNTERFACTUAL ANALYSES

This section studies various counterfactuals to shed light on the equilibrium role and value of application approvals. A detailed description of how we recompute equilibria is provided in Appendix D.3 and our counterfactual results are summarized in Table III. Appendix Figure A7 plots, for each of our counterfactuals, the relationship between search and interest rates, default, approval, as well as the distribution of realized search and rates, and the adverse selection plot showing the share of high types as a function of origination rates.

6.1. *Tighter Lending Standards*

Tightening lending standards has been at the heart of policy debates for many years and has arisen again during the recent COVID-19 pandemic. The debate has frequently centered around the challenge of providing consumers access to credit while mitigating systematic risks in the financial sector (Dell'Ariccia et al. (2012); Mian and Sufi (2009); Bassett et al. (2014)). Famously, Ben Bernanke was declined for a mortgage during his tenure as chairman of the Federal Reserve. Traditionally, a tightening in credit standards is modeled as an increase in the cost of lending, even though mortgage costs may be quite low because of monetary policy. Our model allows us to study the implications of a more realistic scenario, in which there are no changes in the underlying costs of lending but denial rates increase. As we show, tightening lending standards results in higher mortgage rates even if the underlying costs of providing mortgages do not change.

In our model, tightening lending standards reflect reductions in the application approval parameters p_z. To calibrate this drop in p_z, we use our application data to 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, controlling for state

TABLE III

COUNTERFACTUAL SUMMARY

	All Borrowers		High Type		Low Type	
	Average	S.D.	Average	S.D.	Average	S.D.
Realized Interest Rates						
Baseline	-0.002	0.664	-0.384	0.459	0.140	0.673
Tighter Standards	0.252	0.754	-0.217	0.497	0.427	0.759
Redlining – All Borrowers	0.285	0.724	0.243	0.703	0.301	0.732
Redlining – Redlined Group	0.298	0.723				
Redlining – Non-Redlined Group	0.282	0.725				
Search Distribution						
Baseline	3.53	2.67	2.10	1.60	4.11	2.80
Tighter Standards	3.77	2.80	2.25	1.74	4.43	2.92
Redlining – All Borrowers	3.24	2.74	1.12	0.46	4.12	2.81
Redlining – Redlined Group	3.54	2.89				
Redlining – Non-Redlined Group	3.17	2.70				

Supply Effects	Offered Rate Dist.		Bank
	Average	S.D.	Profits
Baseline	0.206	0.723	1.893
Tighter Standards	0.483	0.805	2.130
Redlining – All Lenders	0.300	0.733	1.995
Redlining – Redlining Lenders	0.291	0.753	1.867
Redlining – Non-Redlining Lenders	0.308	0.713	2.124

Notes: Table reports mean and standard deviation of search and realized interest rates across our counterfactual model simulations. Interest rates are residualized against observables in the data; thus, rates should be interpreted as percentage points of return over an observably identical loan in the data. 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. Interest rates and profit margins are expressed in percentage points above the mean realized rate in the market for an observably comparable borrower and loan type. Profits reflect the profits for a bank posting the average realized rate in the market, net of any T1EV profit shocks ξ_{jk}. "Tighter Standards" refers to a counterfactual in which the odds of application approval drop as they did following the recession, by reducing the odds of application approval by 21.8%. "Redlining" supposes that half of the lenders in the market engage in redlining behavior by approving both high and low type borrowers of a discriminated group B, which occupies 20% of the population, at half the rate of a group W.

fixed effects. Our estimates imply a reduction of the odds ratio of approval by 21.8% during the crisis, suggesting that mortgage credit became more difficult to attain for borrowers following the crisis. Our counterfactual mimics this change 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 substantially raises the mean and variance of rates paid by borrowers. The mean rate paid in the market increases by 25.4bp or about 0.6 standard deviations of residualized rates. This increase is on the order of a discrete increment in the Fed's policy rate and corresponds to $301 of higher payments per year for the average loan in our sample. Tightening lending standards also increases the standard deviation of realized interest rates by 9.0bp.

The reason for this large rate response is best understood by considering the borrowers' problem first. As the approval probability p_z falls, borrowers reservation rates rise as in equation (2). That is, all borrowers become more willing to accept a high priced mortgage and so pay higher rates holding fixed the distribution of offered rates. This change in borrower behavior in turn increases the profitability of offering high interest rate loans, incentivizing a lender to offer higher interest rates. In effect, firms' residual demand curves defined by equation (5) become more inelastic. In response to higher offer rates, borrowers' reservation rates increase further. In other words, the amplification from strategic complementarity in rate setting described in section 2.3 is an important driver of the magnitude of our effects.

These patterns are visible graphically in Appendix Figure A7. Interestingly, uniformly tightening credit standards does not affect the extent of adverse selection in this 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. As discussed in section 2, the informativeness of screening technology (p_h relative to p_l), rather than the overall difficulty of obtaining credit (the level of p_z), determines the extent of adverse selection in our model.

Overall, our counterfactual illustrates that even if the costs of lending do not increase during periods of tightened credit standards, the rates paid by borrowers increase. Therefore, since the cost of financing partly reflects the cost of funds, tightening lending standards induces a wedge between mortgage rates and the risk-free rate. The effect of tightened standards on mortgage rates can be substantial – on the order of a discrete increment in the Fed's policy rate. Policies affecting credit standards must account for this effect on realized prices in credit markets, in addition to the standard credit access considerations.

6.2. *Discrimination and Redlining*

Redlining is a practice of discrimination that restricts credit access to consumers based on their socioeconomic, racial, or ethnic makeup. Such policies are increasingly a cause for concern by policymakers who worry that changes in technology alter the screening behavior of new "fintech lenders" with respect to discriminated groups (Fuster et al., 2020). Our model is suited for the analysis of a realistic redlining scenario, in which a portion of lenders in the market discriminate by lowering approval rates for borrowers from the discriminated group. Discrimination of this sort was one of the primary reasons for the establishment of the Home Mortgage Disclosure Act (HMDA), a federal law requiring mortgage lenders to submit records of mortgage applications and rejection decisions to regulators. Such discrimination is more subtle than that in canonical Becker (1957) models – where some lenders do not lend to minorities – or explicitly charging different prices to minority borrowers.[39]

In this counterfactual, suppose 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.[40] A redlining lender approves the discriminated B borrowers at half the rate that the non-discriminated W borrowers of the same creditworthiness are accepted: that is, $p_z^B = 0.5p_z^W$. Half of lenders in the market redline. Non-redlining lenders ignore the B, W distinction. Lenders 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 most consistent with the type of redlining and discrimination which have been alleged in this market.[41] Last, we assume that borrowers are only aware of the proportion of lenders redlining, but not *which* lenders redline. This is consistent with the fact that discriminated borrowers keep applying for loans from lenders which are later alleged to have discriminated.

Despite the absence of discriminatory pricing, discriminated borrowers pay $16bp$ higher rates in equilibrium on average than the non-discriminated W borrowers with the same search cost and creditworthiness. Discriminated borrowers understand that their chances of obtaining a loan approval in the future are low, so they are more willing to accept higher mortgage rates and thus endogenously sort to lenders which offer higher rates. Discriminated borrowers may therefore appear less financially-sophisticated even though their underlying ability to search for mortgages

---

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

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

[41]See, for example, Ladd (1998), and this article by Bloomberg [accessed Jan 17, 2019].

is the same as that of the majority. Discriminated minority borrowers may accept worse mortgages than non-minorities as a rational response to perceiving higher rejection rates. Interpreting data on interest rates and rejections across potentially discriminated groups is therefore only possible in a model which accounts for search behavior and rejections.

Redlining lenders are less profitable, consistent with the intuition of Becker (1957). The primary driver of lost profits are lower volumes, rather than lower prices. Redlining lenders offer $1.6bp$ lower rates than do non-redlining lenders on average. This exacerbates the difference in realized rates between B and W borrowers since B borrowers are less likely to receive a loan from a redlining lender. The small difference in offered rates arises 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 a lender's acceptance probability does not drastically affect the firm's pricing decision.

Due to the strategic complementarities in rate setting, redlining increases the overall interest rates charged by lenders in the market, thereby also hurting the non-discriminated group. Intuitively, rejections of redlined consumers increase their willingness to accept high rates, increasing lenders' incentives to raise 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.

Furthermore, Panel F of Appendix Figure A7 shows that the gradient of the adverse selection curve – the relationship between the share of high types and realized interest rates – is significantly flatter in this redlining counterfactual. This occurs because redlining compresses the gap between high and low type approval probabilities: a subset of high type borrowers are rejected at a high rate through redlining and so are willing to accept high interest rates. Thus, more high type borrowers realize high interest rates in equilibrium. This is another reason why redlining allows banks to offer higher interest rates to the market. Not only does the decline in approval probabilities lead to increases in borrower reservation rates, but the reduction in expected repayments as one increases interest rates is also less pronounced. Both of these forces increase the expected profits from offering high interest rates, leading to higher rates in the market for all borrowers, regardless of whether the borrower belongs to the redlined group or not.

The increase in offered rates does not offset the lost market share for the redlining lenders: their profits profits decline slightly, by 2.6bp, compared with an increase in profits of 23.1bp for the non-redlining lenders, relative to the baseline estimates. Put differently, redlining lenders lose 25.7bp in rate of return relative to their competitors that do not redline.

7. CONCLUSION

Our paper highlights how the presence of rejections changes the conclusions that researchers can draw about search from the data. Ignoring rejections biases search cost estimates, which may lead researchers and policymakers to misclassify the forces responsible for credit allocation, especially for less creditworthy consumers. For example, uncreditworthy borrowers may be classified as having high search cost, when they are instead rationally responding to increased rejection rates. Combining search and credit approval allows us to analyze the equilibrium consequences of policies and innovations which change the approval process, such as restrictions on the information that lenders can use to screen or changes in screening technologies. Many policies have this flavor, such as the Community Reinvestment Act (CRA) and the establishment of mortgage insurance and the FHA. We analyze an illustrative set of counterfactuals and show that accounting for search and rejection simultaneously is critical to understanding the impact of such policies.

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. Fortunately, as we show, the distribution of search costs and approval rates can be estimated when search and pricing outcomes are observed.

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 larger effects on equilibrium prices if they improve both borrowers' sophistication (search cost) and 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 valuable. Finally, building models which explicitly acknowledge the role of soft search in these markets is an important next step in this agenda.

REFERENCES

Agarwal, S., S. Chomsisengphet, N. Mahoney, and J. Stroebel (2018). Do Banks Pass Through Credit Expansions to Consumers Who Want To Borrow? Quarterly Journal of Economics 133(1). [1, 6]

Alexandrov, A. and S. Koulayev (2017). No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information. Consumer Finance Protection Bureau Office of Research Working Paper 01, 1–57. [5, 22]

Allen, J., R. Clark, and J.-F. Houde (2014). The Effect of Mergers in Search Markets: Evidence from the Canadian Mortgage Industry. American Economic Review 104(10), 3365–3396. [1, 2, 5, 22, 35]

Allen, J., R. Clark, and J.-F. Houde (2019). Search Frictions and Market Power in Negotiated-Price Markets. Journal of Political Economy 127(4), 1550–1598. [5, 35]

Ambokar, S. and K. Samaee (2019a). Inaction, Search Costs, and Market Power in the US Mortgage Market. University of Pennsylvania mimeo. [5]

Ambokar, S. and K. Samaee (2019b). Mortgage Search Heterogeneity, Statistical Discrimination and Monetary Policy Transmission to Consumption. University of Pennsylvania mimeo. [5, 19, 22]

Andersen, S., J. Y. Campbell, K. M. Nielsen, and T. Ramadorai (2015). Inattention and Inertia in Household Finance: Evidence from the Danish Mortgage Market. NBER Working Paper Series 21386. [5]

Argyle, B., T. Nadauld, and C. J. Palmer (2017). Real Effects of Search Frictions in Consumer Credit Markets. MIT Sloan Working Paper (5242-17). [1, 6]

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

Bassett, W. F., M. B. Chosak, J. C. Driscoll, and E. Zakrajsek (2014). Changes in bank lending standards and the macroeconomy. Journal of Monetary Economics 62, 23–40. [37]

Becker, G. (1957). The Economics of Discrimination (2 ed.). The University of Chicago Press. [40, 41]

Benneton, M. (2019). Leverage Regulation and Market Structure: A Structural Model of the UK Mortgage Market. Working Paper. [6]

Bhutta, N., A. Fuster, and A. Hizmo (2021). Paying Too Much? Borrower Sophistication and Overpayment in the US Mortgage Market. Working Paper. [2, 22]

Buchak, G., G. Matvos, T. Piskorski, and A. Seru (2020). Beyond the Balance Sheet Model of Banking: Implications for Bank Regulation and Monetary Policy. Working Paper. [6]

Buchak, G., T. Piskorski, and A. Seru (2018). Fintech, Regulatory Arbitrage, and the Rise of Shadow Banks. Journal of Financial Economics 130(3), 453–483. [6]

Calem, P. S. and L. J. Mester (1995). Consumer Behavior and the Stickiness of Credit-Card Interest Rates. American Economic Review 85(5), 1327–1336. [1]

Carlson, J. A. and R. P. McAfee (1983). Discrete Equilibrium Price Dispersion. Journal of Political Economy 91(3), 480–493. [1, 7, 9]

Dell'Ariccia, G., D. Igan, and L. Laeven (2012). Credit Booms and Lending Standards: Evidence from the Subprime Mortgage Market. Journal of Money Credit and Banking 44(2), 367–384. [37]

Fuster, A., P. Goldsmith-Pinkham, T. Ramodorai, and A. Walther (2020). Predictably unequal? The effects of machine learning on credit markets. Working Paper. [39]

Gilbukh, S. and P. Goldsmith-Pinkham (2020). Heterogeneous Real Estate Agents and the Housing Cycle. Working Paper. [6]

Guerrieri, V., R. Shimer, and R. Wright (2010). Adverse Selection in Competitive Search Equilibrium. Econometrica 78(6), 1823–1862. [5]

Gurun, U. G., G. Matvos, and A. Seru (2016). Advertising Expensive Mortgages. Journal of Finance LXXI(5), 2371–2416. [1, 5, 22, 24]

Handel, B., I. Hendel, and M. D. Whinston (2015). Equilibria in Health Exchanges. Econometrica 83(4), 1261–1313. [5, 10]

Hortaçsu, A. and Syverson (2004). Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P 500 Index Funds. Quarterly Journal of Economics 119(2), 403–456. [2, 4, 14, 35]

Jarosch, G. and L. Pilossoph (2019). Statistical Discrimination and Duration Dependence in the Job Finding Rate. Review of Economic Studies 86(4), 1631–1665. [18]

Koijen, R. and M. Yogo (2016). Shadow Insurance. Econometrica 84(3), 1265–1287. [6]

Koijen, R. and M. Yogo (2020). The Fragility of Market Risk Insurance. Working Paper. [6]

Ladd, H. F. (1998). Evidence on Discrimination in Mortgage Lending. Journal of Economic Perspectives 12(2), 41–62. [40]

Lang, K., M. Manove, and W. T. Dickens (2005). Racial Discrimination in Labor Markets with Posted Wage Offers. American Economic Review 95(4), 1327–1340. [40]

Lester, B., A. Shourideh, V. Venkateswaran, and A. Zetlin-Jones (2016). Screening and Adverse Selection in Frictional Markets. Federal Reserve Bank of Philadelphia Working Paper Series 16. [5]

McCall, J. J. (1970). Economics of Information and Job Search. Quarterly Journal of Economics 84(1), 113–126. [1, 7]

Mian, A. and A. Sufi (2009). The Consequences of Mortgage Credit Expansion: Evidence from the U.S. Mortgage Default Crisis. Quarterly Journal of Economics 124(4), 1449–1496. [4, 37]

Piazzesi, M., M. Schneider, and J. Stroebel (2020). Segmented Housing Search. American Economic Review 110(3). [6]

Riley, J. G. (1979). Informational Equilibrium. Econometrica 47(2). [10]

Rothschild, M. (1974). Searching for the lowest price when the distribution of prices is unknown. Journal of Political Economy 82, 689–711. [21]

Roussanov, N., H. Ruan, and Y. Wei (2018). Marketing Mutual Funds. NBER Working Paper Series 25056. [2]

Salz, T. (2017). Intermediation and Competition in Search Markets: An Empirical Case Study. Working Paper. [11]

Scharfstein, D. and A. Sunderam (2017). Market Power in Mortgage Lending and the Transmission of Monetary Policy. Working Paper. [6]

Stango, V. and J. Zinman (2013). Borrowing High vs Borrowing Higher: Sources and Consequences of Dispersion in Individual Borrowing Costs. NBER Working Paper Series 19069. [1]

Stigler, G. J. (1961). The Economics of Information. Journal of Political Economy 69(3). [7, 67]

Stroebel, J. (2015). Asymmetric Information about Collateral Values. Journal of Finance 71(3). [1, 4]

Wong, A. (2019). Refinancing and the Transmision of Monetary Policy to Consumption. Working Paper. [6]

Woodward, S. E. and R. E. Hall (2012). Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence. American Economic Review 102(7), 3249–3276. [1, 5, 23]