Econometrica Supplementary Material

SUPPLEMENT TO "SEARCHING FOR APPROVAL"

(*Econometrica*, Vol. 92, No. 4, July 2024, 1195–1231)

SUMIT AGARWAL
Departments of Economics, Finance, and Real Estate, National University of Singapore

JOHN GRIGSBY
Department of Economics, Princeton University and NBER

ALI HORTAÇSU
Kenneth C. Griffin Department of Economics, University of Chicago and NBER

GREGOR MATVOS
Kellogg School of Management, Northwestern University and NBER

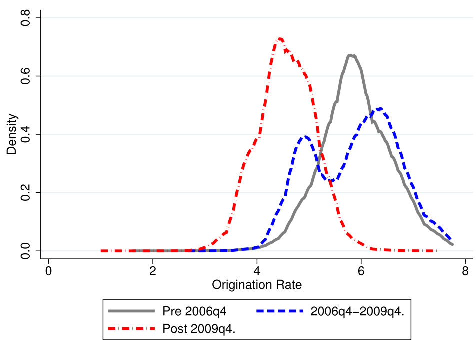
AMIT SERU
Graduate School of Business, Stanford University and NBER

VINCENT YAO
Department of Real Estate, Georgia State University

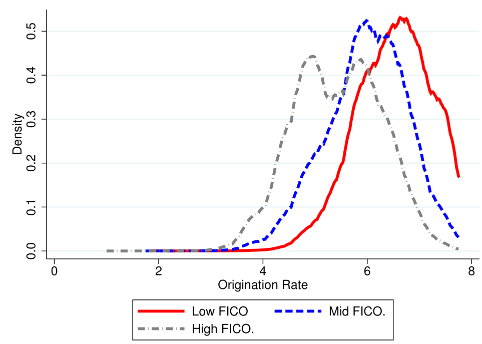
APPENDIX A INCLUDES ADDITIONAL TABLES AND FIGURES SHOWING the robustness of our core results. Appendix B presents evidence that inquiries are the best available proxy of search in the mortgage market using data from HMDA and the NSMO. Appendix C describes the additional robustness figures of Appendix A. Appendix D presents additional details on the estimation and simulation of the model, including a detailed derivation of the likelihood function. Appendix E explores the robustness of the theoretical results if search is simultaneous rather than sequential.

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

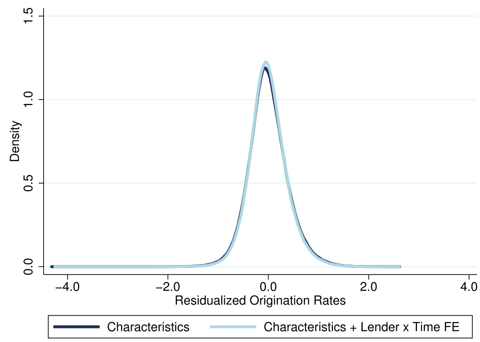
APPENDIX A: ADDITIONAL ROBUSTNESS TABLES AND FIGURES



Panel A: Raw Rates by Origination Date

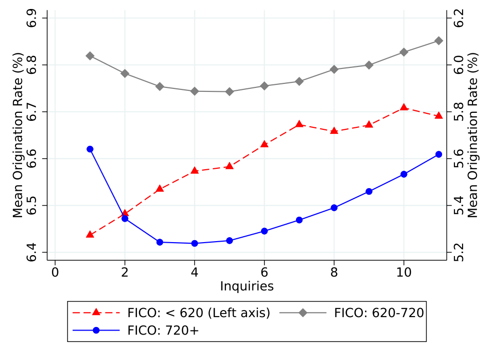


Panel B: Raw Rates by Borrower FICO Score

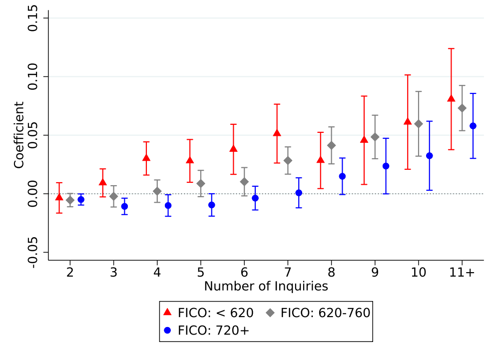


Panel C: Rates Residualized Against Observables

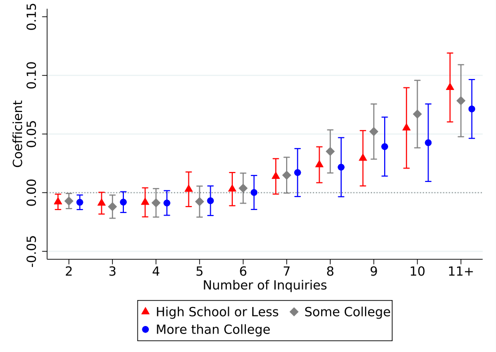
FIGURE A1.—Distribution of mortgage rates in the U.S. *Notes*: Figure plots the kernel-density estimated distribution of mortgage rates in the U.S. using our loan data. Panel A plots the raw observed rates across three time periods: before the house price peak of September 2006, between the house price peak and end of the crisis in 2009, and the post-crisis period from the first quarter of 2010 on. Panel B plots the distribution of observed mortgage rates for three borrower FICO buckets: low FICO ($\leq$620), middle FICO (620–719), and high FICO (720+). Finally, Panel C plots the distribution of residuals from a regression of realized interest rates on borrower and loan characteristics. The black line residualizes against only borrower characteristics, which include the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter characteristics. The light blue line plots residuals from a regression of rates on these borrower characteristics as well as lender $\times$ origination quarter fixed effects.



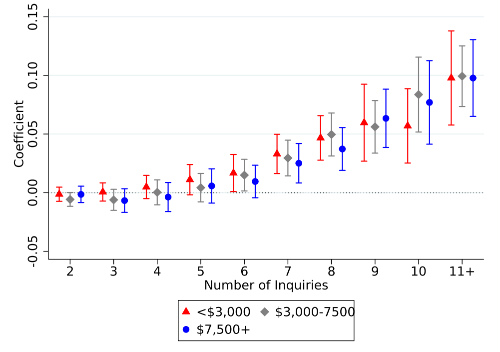
Panel A: Rates by FICO, Raw



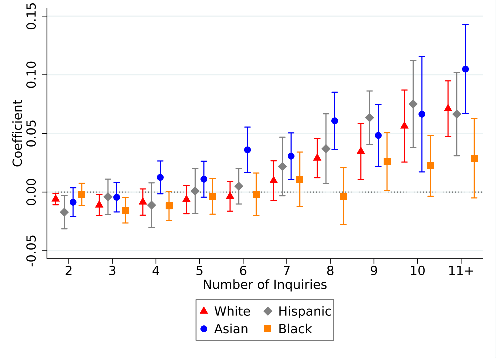
Panel B: By FICO, Cond. on Covariates



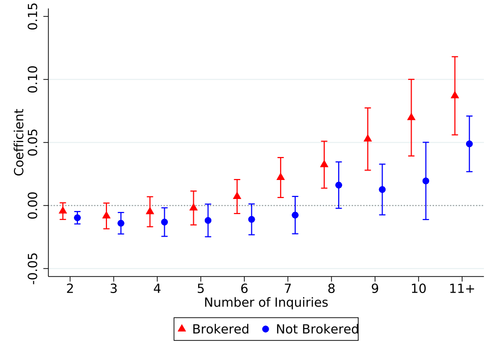
Panel C: By Education



Panel D: By Monthly Income

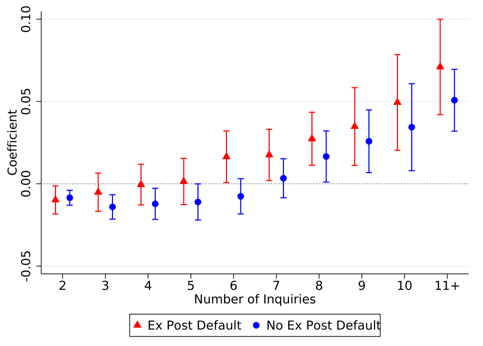


Panel E: By Race

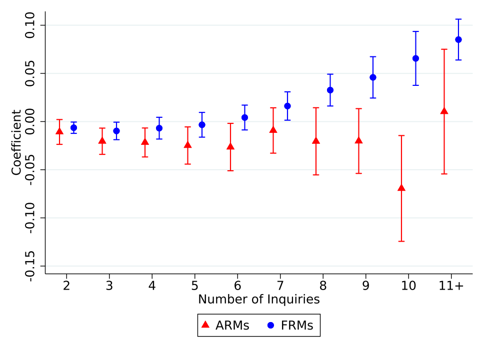


Panel F: By Broker Status

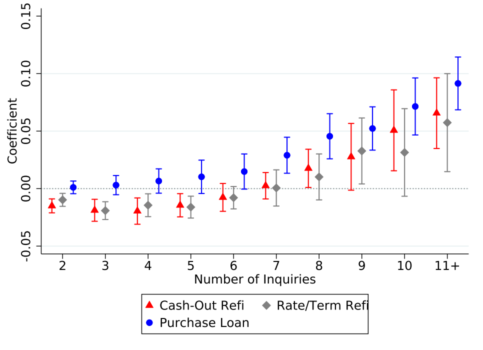
FIGURE A2.—Robustness of relationship between inquiries and interest rates: borrower characteristics. *Notes*: Figure plots average realized interest rates against inquiry counts in our loan data. Panel A plots the unconditional relationship across three FICO buckets: low FICO ($\leq$620), middle FICO (620–719), and high FICO (720+). Panels B through F plot regression coefficients estimated from equation (10) using OLS on subsamples defined by FICO buckets (Panel B), education (Panel C), monthly income (Panel D), race (Panel E), and whether the borrower used a broker to originate the loan (Panel F). 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 1 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. 95% confidence intervals reported on plot.



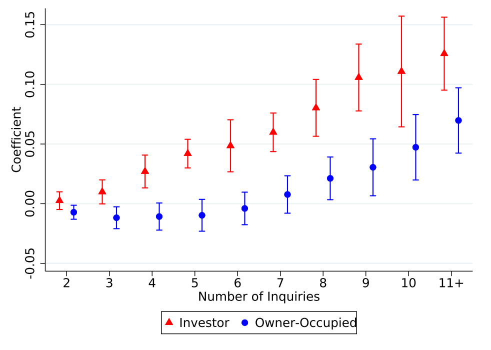
Panel A: By Ex Post Default



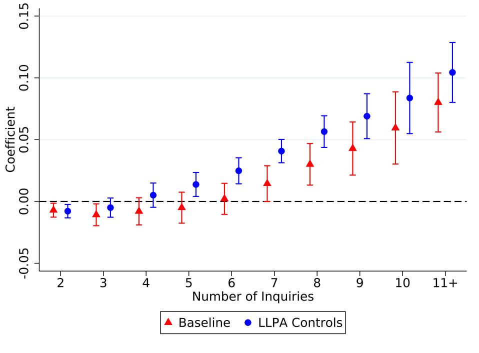
Panel B: By Product Type



Panel C: By Loan Purpose

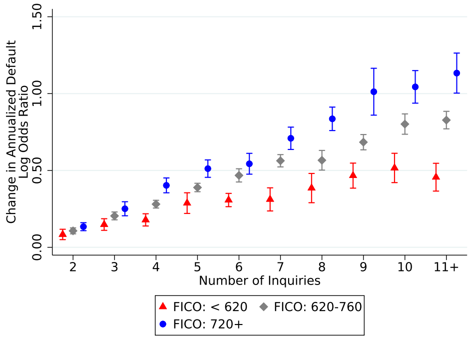


Panel D: By Investor Status

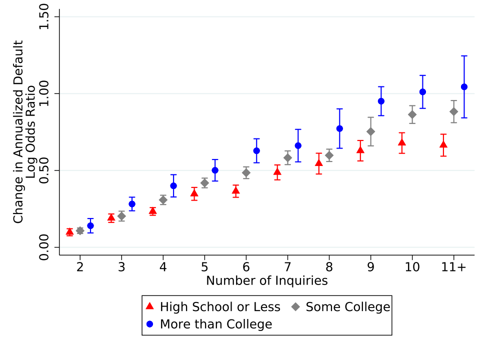


Panel E: Controlling for LLPA Controls

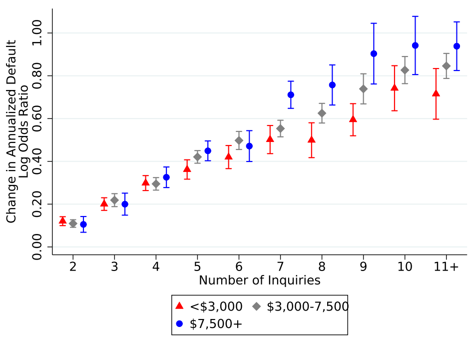
FIGURE A3.—Relationship between search and realized mortgage interest rates, conditional on observables, by ex post delinquency status and brokerage status. *Notes*: Figure plots regression coefficients estimated from equation (10) using OLS for the separate subsamples of loans which do/do not default ex post (Panel A), for ARMs versus FRMs (Panel B), for refinance and purchase loans (Panel C), for loans originated by an investor or owner-occupant (Panel D), and adding additional controls (Panel E). The dependent variable in each regression is the origination interest rate on a loan plus points and fees. Default defined as a loan being in foreclosure or at least 90 days delinquent by Jan 1, 2015. The independent variables are a set of dummy variables equal to 1 if the inquiry count at mortgage origination were equal to s for s in $\{2, 3, 4, \dots, 11+\}$. 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. Panel E additionally controls for Loan-Level Price-Adjustment (LLPA) categories from Fannie Mae. 95% confidence intervals reported on plot. Figure uses our loan data.



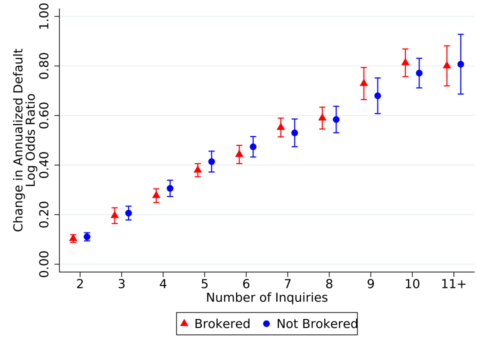
Panel A: By FICO



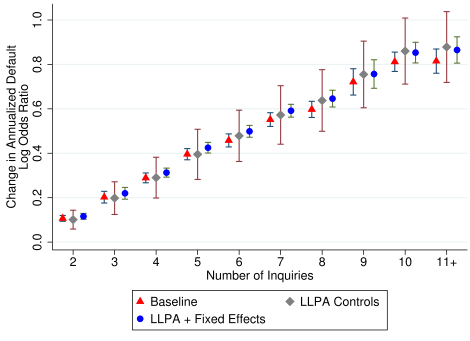
Panel B: By Education



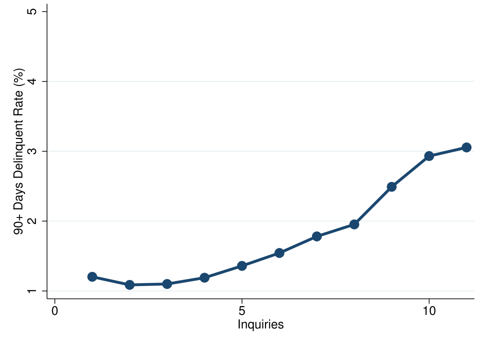
Panel C: By Monthly Income



Panel D: By Broker Status

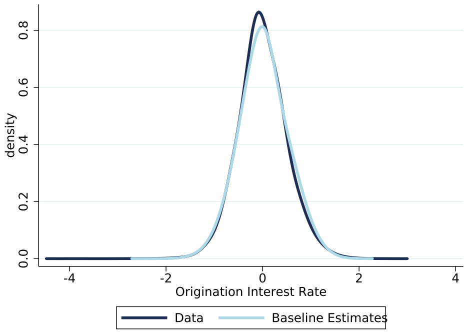


Panel E: Controlling for LLPA Categories

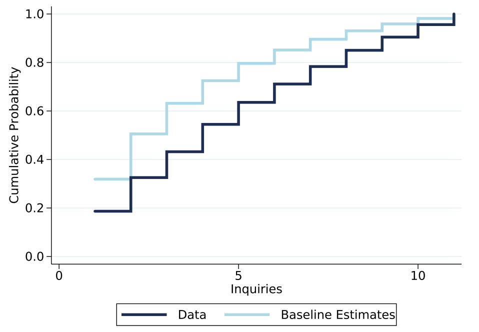


Panel F: 90 Day Delinquency

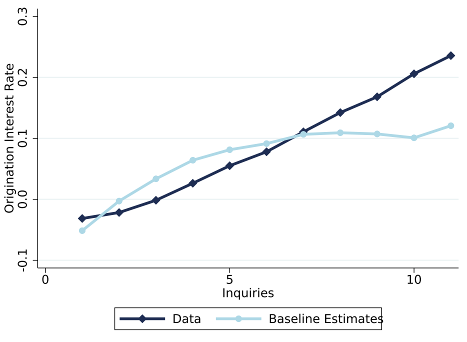
FIGURE A4.—Robustness of relationship between inquiries and default rates. *Notes*: Figure shows robustness of the positive relationship between default rates and inquiry counts. Panels A through E: regression coefficients estimated from equation (11) for subsamples of borrowers defined by FICO score (Panel A), education (Panel B), monthly income (Panel C), Broker status (Panel D), or controlling for LLPA categories (Panel E). The coefficients reflect changes in the log odds ratio of the annual default hazard relative to borrowers with one inquiry. These panels define default as the loan being at least 90 days delinquent, or entering foreclosure. Panel F defines default as a loan being at least 90 days delinquent. The independent variables are a set of dummy variables equal to 1 if the inquiry count at mortgage origination equals s for s in $\{2, 3, 4, \dots, 11+\}$. The omitted category is $s = 1$. Controls are included for the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination quarter. Standard errors are clustered at the origination quarter level. Panel E additionally controls for Loan-Level Price-Adjustment (LLPA) categories from Fannie Mae. 95% confidence intervals reported on plot. Figure uses our loan data.



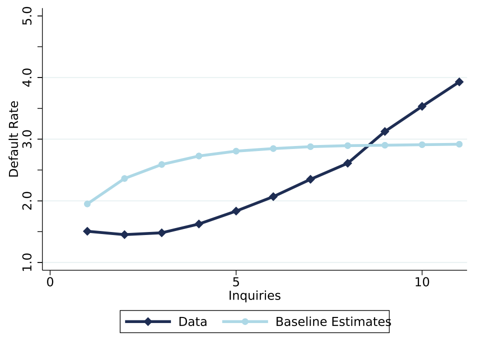
Panel A: Realized Residual Rate Densities



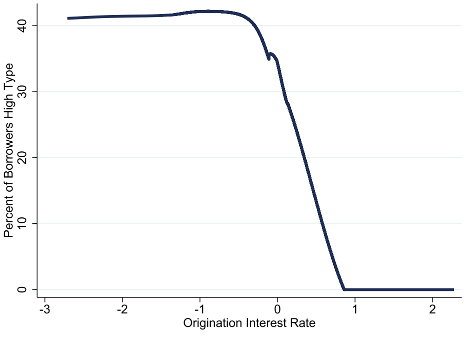
Panel B: Search Distribution



Panel C: Search and Origination Rates

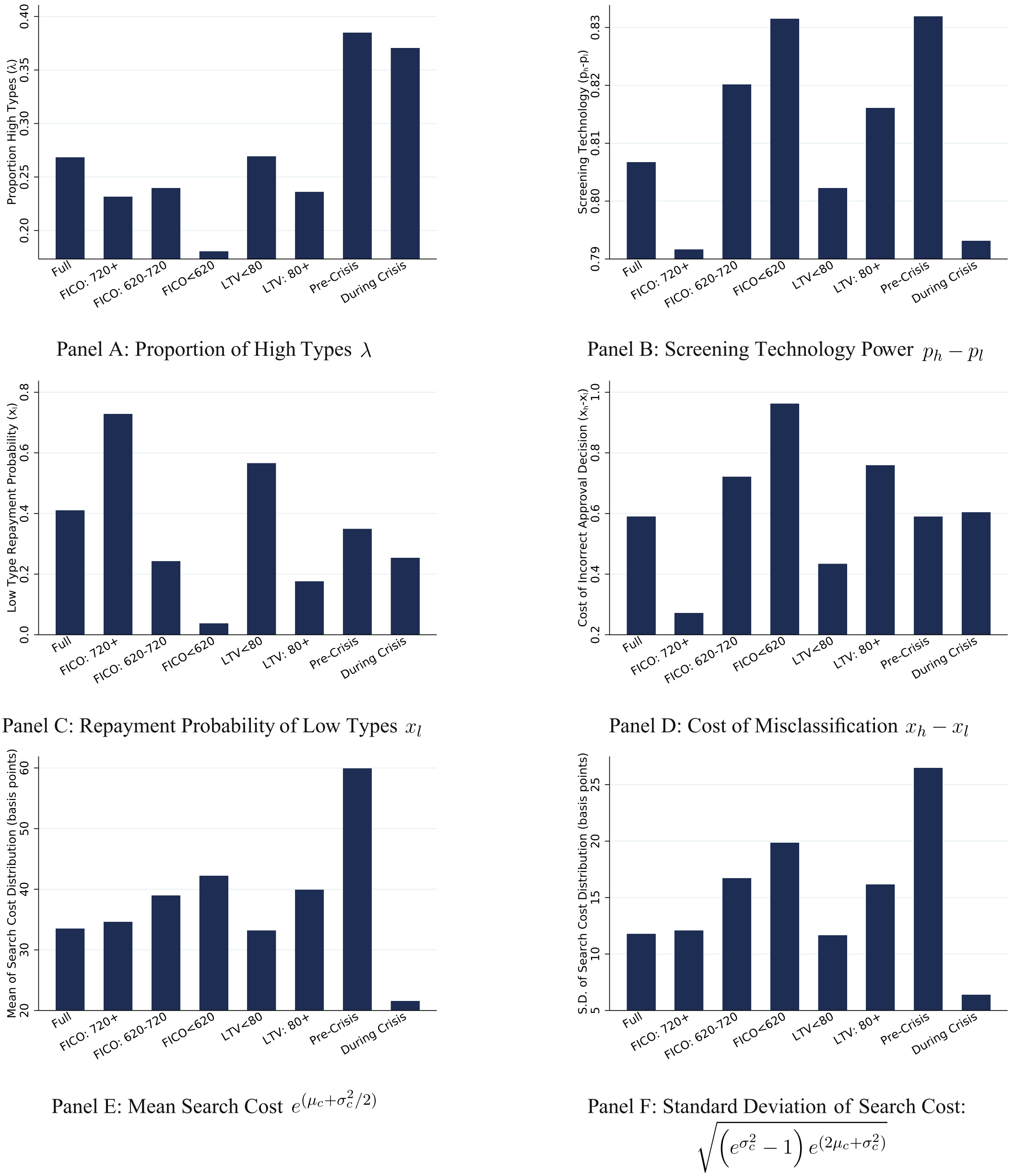


Panel D: Search and Default Probability



Panel E: Share of High Types as Function of Origination Rate

FIGURE A5.—Model performance: search behavior in data versus model simulation with estimated parameters. *Notes*: Figure plots the performance of our model under our benchmark estimated parameters from Table II. Black lines plot quantities in our estimation sample, while light blue lines plot those implied by a large model simulation using parameters estimated by maximum likelihood following the approach laid out in Section 5.1. Origination rates in data residualized against the borrower's FICO score, combined loan-to-value (LTV) ratio, backend debt-to-income (DTI) ratio, refinance and product type indicators, state fixed effects, and origination year fixed effects. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panels C and D show the relationship between search and origination interest rates and default probability, respectively, where default probability is measured as of January 2015. To compute these default probabilities in the simulation, we randomly draw a mortgage's origination date from the distribution of origination dates in the data. Panel E shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate r who are of high type. Estimation uses both loan and application data sets.



Panel A: Proportion of High Types λ

Panel B: Screening Technology Power $p_h - p_l$

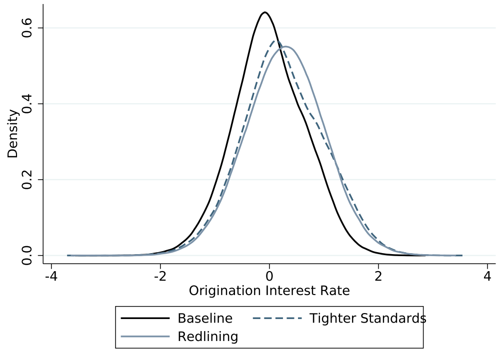
Panel C: Repayment Probability of Low Types x_l

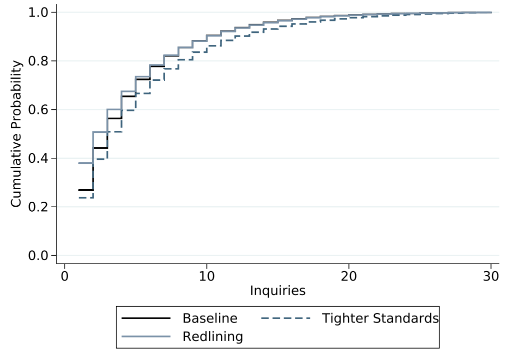
Panel D: Cost of Misclassification $x_h - x_l$

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

Panel F: Standard Deviation of Search Cost: $\sqrt{\left(e^{\sigma_c^2} - 1\right) e^{(2\mu_c + \sigma_c^2)}}$

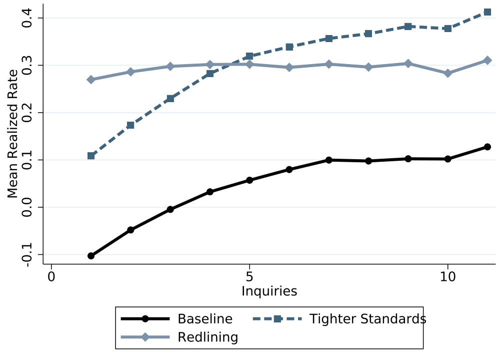
FIGURE A6.—Estimates by subsample. *Notes*: Figure shows estimated parameter values from our maximum likelihood routine across eight subsamples. The sample of borrowers originating their mortgage in 2010 or later is omitted due to small sample size. The acceptance probability for high types p_h is 1 for all subsamples. Estimation uses both loan and application data sets.

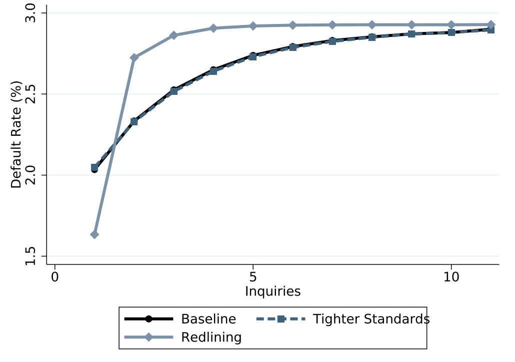




Panel A: Realized Rate Distribution

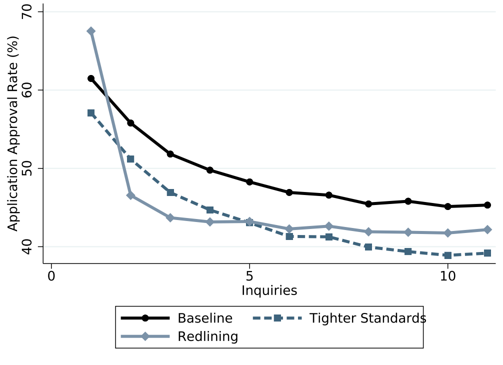
Panel B: Distribution of Search

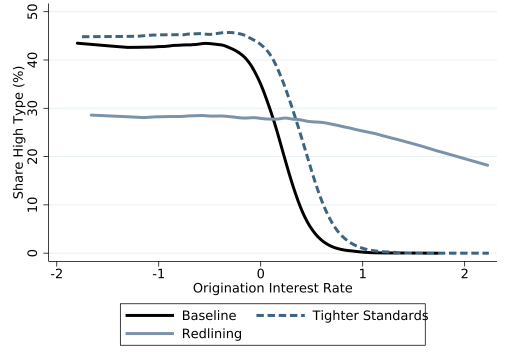




Panel C: Relationship Between Search and Price

Panel D: Relationship Between Search and Default





Panel E: Relationship Between Search and Application Approval

Panel F: Share of High Types as Function of Origination Rate

FIGURE A7.—Counterfactual plots. *Notes*: Figure plots the key aspects of search behavior under our baseline parameter estimates (black line) and our three counterfactuals. The dashed navy line reports the Tighter Lending Standards Counterfactual where the odds of application acceptance decline by 21.8% for both high and low type borrowers. The solid medium blue line reports the redlining counterfactual. Panel A plots the density of realized interest rates in the market. Panel B plots the CDF of search for realized loans. Panel C, D, and E show the relationship between search and origination interest rates, the probability of ever defaulting, and the application approval rate. Panel F shows the degree of adverse selection in the market by plotting the share of all borrowers originating a mortgage at rate r who are of high type.

TABLE A1

AVERAGE BORROWER AND LOAN CHARACTERISTICS BY TIME PERIOD.

Origination Date Relative to 2006q4–2009q4:	Loan Data			Application Data		
	Pre	During	Post	Pre	During	Post
Search and Rates						
# Inquiries	1.87	3.16	3.29	6.00	5.28	4.95
Pr{Approval} (%)	–	–	–	85.92	87.26	87.22
Origination Interest Rate (%)	5.91	5.87	4.56	–	–	–
Creditworthiness						
FICO	713.0	726.3	762.1	712.6	721.3	753.2
CLTV	74.1	75.5	68.6	73.2	74.6	72.3
Backend DTI ratio	36.98	39.63	33.92	36.50	38.23	32.84
Pr{Default} (Annualized %)	2.05	3.06	0.31	–	–	–
Pr{90+ Days Delinquent} (%)	1.24	1.82	0.25	–	–	–
Loan Characteristics						
FRM 30-year (%)	71.7	85.5	66.1	67.8	83.0	68.0
FRM 15-year (%)	23.5	12.7	27.7	22.4	11.8	27.1
ARM (%)	4.78	1.83	6.20	9.84	5.15	4.84
Loan Origination Amount ($ 000s)	138.4	187.3	210.6	–	–	–
Cash-out refi (%)	33.7	30.2	25.5	–	–	–
Rate-term refi (%)	26.7	25.0	41.2	–	–	–
Borrower Characteristics						
White (%)	80.4	77.4	80.2	–	–	–
Black (%)	8.5	8.1	3.0	–	–	–
Borrower Male (%)	44.5	42.0	40.9	–	–	–
Borrower Age	43.6	44.3	47.4	–	–	–
Less than High School (%)	26.4	27.9	18.6	–	–	–
High School and Some College (%)	46.5	53.8	55.3	–	–	–
College or more (%)	16.1	17.9	26.1	–	–	–
Borrower Monthly Income ($)	5088	6463	8095	–	–	–
Investor (%)	7.2	9.1	10.9	6.0	7.5	8.5
Observations (000s)	574	549	194	1326	899	1039

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

TABLE A2

PREDICTORS OF INQUIRY COUNTS AMONG MORTGAGE APPLICANTS.

	# Inquiries (1)	1^{st} Quartile (2)	2^{nd} Quartile (3)	3^{rd} Quartile (4)	4^{th} Quartile (5)
			Panel A: Loan Data		
FICO score (std)	−0.389	6.570	1.192	−0.665	−7.096
	(0.030)	(0.271)	(0.409)	(0.226)	(0.604)
Combined LTV (std)	0.099	−2.261	0.016	0.371	1.874
	(0.007)	(0.129)	(0.147)	(0.086)	(0.160)
Back-end DTI Ratio (std)	0.120	−2.247	−0.316	0.235	2.328
	(0.011)	(0.096)	(0.168)	(0.076)	(0.256)
FRM 15-year	−0.271	5.266	0.777	−0.887	−5.156
	(0.024)	(0.390)	(0.556)	(0.225)	(0.552)
FRM 30-year	−0.157	3.863	−0.197	−0.892	−2.774
	(0.015)	(0.576)	(0.398)	(0.208)	(0.288)
Cash-out refi	−0.141	1.261	1.561	0.327	−3.149
	(0.040)	(0.690)	(0.283)	(0.163)	(0.842)
Black	0.270	−4.097	−0.667	0.147	4.616
	(0.019)	(0.385)	(0.240)	(0.223)	(0.330)
College	−0.109	1.838	0.289	−0.193	−1.934
	(0.014)	(0.287)	(0.111)	(0.109)	(0.257)
Monthly Income < \$3000	−0.173	3.534	0.273	−0.440	−3.368
	(0.017)	(0.138)	(0.305)	(0.135)	(0.395)
Investor	0.456	−6.284	−1.687	−0.163	8.133
	(0.018)	(0.575)	(0.409)	(0.302)	(0.442)
Observations	1,023,931	1,023,931	1,023,931	1,023,931	1,023,931
R-squared	0.2378	0.2232	0.0100	0.0260	0.1731
			Panel B: Application Data		
FICO score (std)	−3.881	9.256	4.831	−1.345	−12.743
	(0.091)	(0.178)	(0.289)	(0.229)	(0.332)
Combined LTV (std)	0.838	−2.853	−0.742	0.909	2.686
	(0.063)	(0.078)	(0.149)	(0.058)	(0.231)
Back-end DTI Ratio (std)	0.555	−2.255	−0.474	0.914	1.815
	(0.018)	(0.133)	(0.086)	(0.084)	(0.072)
FRM 15-year	−1.899	6.303	1.950	−1.940	−6.314
	(0.058)	(0.382)	(0.311)	(0.264)	(0.235)
FRM 30-year	−1.404	4.038	1.608	−0.932	−4.714
	(0.058)	(0.194)	(0.178)	(0.153)	(0.202)
Cash-out refi	−1.045	1.099	2.043	0.682	−3.825
	(0.132)	(0.593)	(0.407)	(0.444)	(0.444)
Investor	3.048	−7.640	−3.796	1.516	9.920
	(0.142)	(0.370)	(0.364)	(0.379)	(0.458)
Observations	5,202,721	5,202,721	5,202,721	5,202,721	5,202,721
R^2	0.2096	0.1106	0.0190	0.0089	0.1558

Note: Estimated coefficients from regression equation (9) reported. Panel A reports estimates for the sample of mortgage applications, while Panel B reports estimates for the sample of realized mortgage borrowers. Column 1 reports coefficients from a regression in which the dependent variable is the number of inquiries on an applicant's credit report. Columns 2 through 5 report coefficients from a regression in which the dependent variable is an indicator variable, scaled by 100, for whether the applicant was in the first, second, third, or fourth quartile of inquiries, respectively. Variables labeled "std" have been standardized to have zero mean and unit standard deviation. Standard errors clustered at the origination quarter level reported in parentheses beneath coefficient. All regressions include origination quarter × state fixed effects.

APPENDIX B: INQUIRIES AS A PROXY FOR SEARCH

This section validates 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 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 an alternative data source for search in the mortgage market—the National Survey of Mortgage Originations (NSMO)—and find that there is likely ambiguity in the appropriate measure of search in the survey data.

B.1. *Benchmarking to Credit Bureau and HMDA Data*

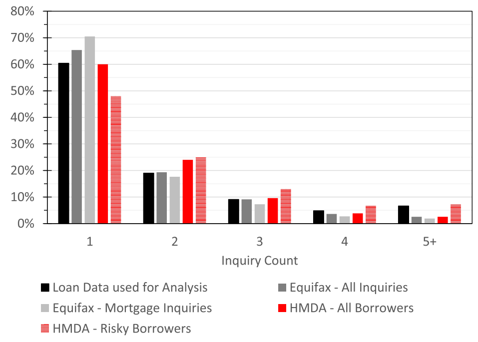
Panel A of Figure B1 compares the distribution of inquiries in our loan data against alternative data sets. The black bars plot the inquiry distribution within our data set of loans used for analysis. The dark and light gray bars plot the distribution of total and mortgage-related inquiries, respectively, in Equifax among people who have changes in mortgage debt. The distributions are very similar, suggesting that neither selection into our sample nor a focus on total inquiries rather than mortgage-inquiries substantially distorts our distribution of inquiries.

Next, we turn to data from the Home Mortgage Disclosure Act (HMDA) to show how this inquiry distribution may emerge from empirical search and rejection patterns. HMDA was enacted in 1975 in order to limit housing discrimination. Under the law, all mortgage lenders are required to report the universe of mortgage applications they receive. Crucially, the data contain information on whether the application was rejected, originated, or whether the borrower withdrew their application. In addition, the data contain some information on loan and borrower characteristics, such as the borrower's race and sex. Starting in 2018, the data also report the interest rate of originated loans and additional characteristics of the application, such as a range for the loan-to-value ratio and debt-to-income ratio.

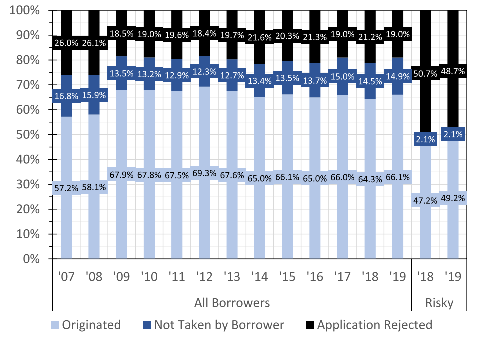
Panel B of Figure B1 plots the share of applications that are rejected, originated, or withdrawn by the borrower in the HMDA data over time. The HMDA data show that between 18% and 26% of applications are rejected in a given year, and an additional 12–17% of applications are withdrawn by the borrower. As a result, only 57–69% of applications eventually originate to loans.[1] Application rejection probability is higher during the house price collapse in 2007 and 2008, which motivates our counterfactual studying tightened lending standards.

We use these patterns to conduct a simple back-of-the-envelope calculation to validate the observed distribution of inquiries. The outcome of this calculation is represented by the red bars in Panel A. For simplicity, suppose that search only occurs if a loan application is not originated and each application generates an inquiry in our data. This is, of course, a lower bound on search and inquiries, as borrowers may inquire about rates and choose not to apply. Furthermore, assume that each application has an i.i.d. chance

[1] We define "loan origination" to be loans with codes that are either "Loan originated" or "Loan purchased by the institution." Applications which are withdrawn or not taken by the borrower are defined by the codes "Application approved but not accepted," "Application withdrawn by applicant," and "Preapproval request approved but not accepted." Rejected applications are those which have codes corresponding to "Application denied by financial institution," "File closed for incompleteness," or "Preapproval request denied by financial institution".



Panel A: Implied Inquiry Distributions



Panel B: HMDA Outcomes

FIGURE B1.—Application rejection and origination probability in HMDA by year. *Notes*: Panel A plots the distribution of inquiries in our loan data, in the Equifax Consumer Credit Panel, and implied by a back-of-the-envelope calculation using HMDA data. Panel B plots the probability that a mortgage application is rejected (light blue), not taken up by the borrower (medium blue), and originated (dark blue) by application year using data from HMDA. Starting in 2018, the data also report the interest rate of originated loans and additional characteristics of the application, such as a range for the loan-to-value ratio and debt-to-income ratio. Risky borrowers are defined to be those with debt-to-income (DTI) ratios of at least 50% and loan-to-value (LTV) ratios above 95%.

of being originated with an origination probability of $\tilde{p}$. Under these assumptions, the probability that an individual with an originated loan would search s times is $\tilde{p}(1-\tilde{p})^{s-1}$. Calibrating $\tilde{p}=0.6$ to roughly match the average origination rate in Figure B1, this would imply that 60% of originated mortgages will have 1 inquiry on record, 24% $(0.4*0.6)$ would have 2 inquiries, 9.6% would have 3, 3.8% would have 4, and 2.6% would have 5+ inquiries. Therefore, even if rejections and withdrawn applications were the only source of search, one would expect to see a substantial right tail of search, as we do in our loan data (Figure 2).

We emphasize, however, that this exercise ignores many considerations which would likely fatten the right tail of the search distribution. If borrowers inquire about loans to receive a rate quote but never apply for the loan, such search will be counted as an inquiry but would not show up in the HMDA data. In addition, the assumption that the origination probability is i.i.d. is likely flawed: in our model, borrowers who have been rejected are more likely to be rejected again, as they are likely to be low type. Therefore, one would expect that the distribution of search implied by this naive back-of-the-envelope calculation underweights the true right tail of the distribution.

Borrowers who are observably less creditworthy are more likely to have their applications rejected. For instance, rejection rates amongst borrowers with DTI ratios above 50% and CLTV ratios above 95% are very high, with roughly half of all applications rejected. This group, however, is unlikely to withdraw their application, suggesting a greater willingness to accept any loan for which they are approved. This is consistent with our model—those who have a high rejection probability (low p_z) are more willing to accept any loan for which they are approved.

We repeat the above back-of-the-envelope calculation for this group by setting $\tilde{p}=0.48$. Doing so yields a much thicker right tail of search: under an i.i.d. origination probability and no other search, 48% of borrowers in this group would have 1 inquiry, 25% $(0.48*0.52)$ would have 2 inquiries, 13% would have 3, 7% would have 4, and 7% would have 5+. This exercise illustrates how a distribution of search may arise with a right tail similar to that of our originated loan data set.

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 more willing to originate unattractive loans. Indeed, interest rates are approximately 22 basis points (¼ of a standard deviation) higher in loan applications that are approved but not taken up than among originated loans, after controlling for DTI, LTV, and loan characteristics (e.g., FRM vs. ARM, refi vs. purchase, etc.). As in our model, it appears that two margins generate the search distribution in the data: a substantial rejection probability and an unwillingness amongst high type borrowers to accept high rate loans.

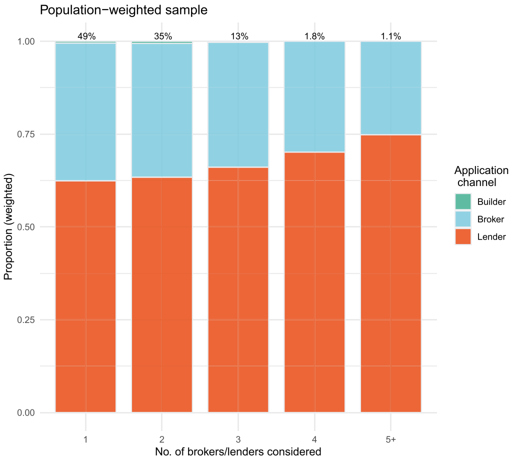
B.2. *Benchmarking to National Survey of Mortgage Originations*

We also seek to reconcile our inquiry-based search metric with those in the National Survey of Mortgage Originations (NSMO), in which 84% of borrowers search no more than twice. The main questions used to measure search in this survey are "How many different lenders/mortgage brokers did you seriously consider before choosing where to apply for this mortgage?" and "How many different mortgage lenders/brokers did you end up applying to?" The wording of these questions offers some explanation for the discrepancy between survey responses and our inquiry data. First, the question combines brokers and lenders into one category, while inquiry data need not. Were a broker to engage in search on a borrower's behalf, they may open multiple inquiries that would only count as one search in the survey data. Since brokers similarly seek a low price for borrowers, it is reasonable to presume that they behave in a similar manner to that described in our model. Indeed, Figure A3 shows that brokered loans are akin to loans originated without a broker.

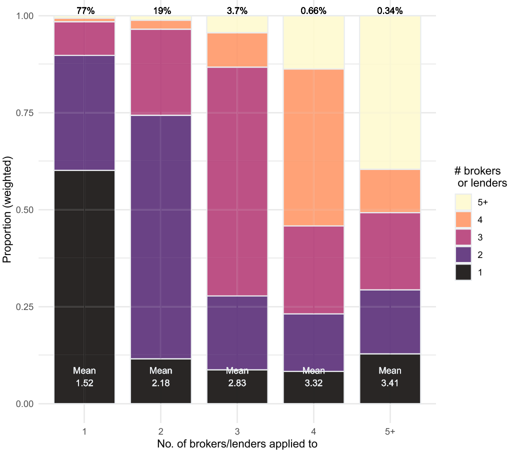
Figure B2 Panel A plots the percentage of borrowers that found their mortgage through a broker against the number of brokers/lenders that a borrower considered. Approximately one-third of borrowers who only considered one broker/lender found their loan through a broker. Only 25% of borrowers who considered five or more brokers/lenders found their loan through a broker. Thus, it is likely that these borrowers with low reported search in the survey may in fact have more "true" search if a broker searches on their behalf. This would be captured by our inquiry data.

Second, it is difficult to know exactly what survey respondents report when asked how many lenders they "seriously considered." For example, survey respondents may not report that they seriously considered rejected applications or mortgages offering high interest rates. However, since they had to pay search costs to find these offers, a search model would consider these as search.[2] One way to see this is to consider the share of borrowers who report applying to more mortgages than they report seriously considering. Panel B of Figure B2 sheds some light on this. The horizontal axis represents the number of lenders/brokers the borrower ends up applying to. Within each category, there is a bar of different colors. Each color corresponds to a number of lenders/brokers seriously considered, with brighter colors representing more lenders considered. For example, about 60% of borrowers who applied to one lender only considered one lender, and around 30% considered two lenders. Remarkably, around 60% of those who apply to five or more

[2]Survey data do not capture soft search well as respondents may not state that they "seriously considered" rates they observe while engaging in soft search activities such as searching the internet for low rate lenders.



Panel A: Probability Originating Through Broker



Panel B: Distribution of Lenders Considered by Number of Lenders Applied to

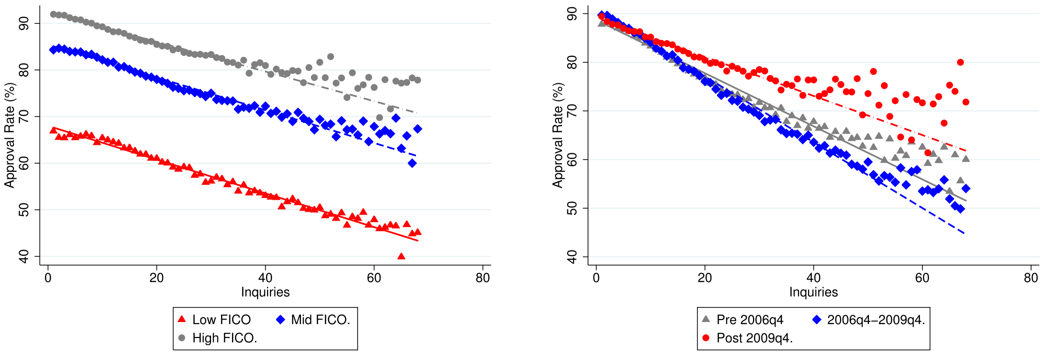
FIGURE B2.—Evaluating search in the National Survey of Mortgage Originations. *Notes*: Figure reports information on search in the National Survey of Mortgage Originations (NSMO). Panel A plots the probability that a borrower originates a mortgage through a broker by the number of brokers/lenders the responder reports "seriously considering." The light blue bar represents the probability of finding the mortgage through a broker, while the darker red bar reports the probability of originating the loan directly from a lender. Panel B reports the distribution of mortgage brokers/lenders seriously considered by the number of brokers/lenders applied to. Darker colors indicate fewer brokers/lenders were considered. For instance, around 60% of borrowers who applied to 1 lender report seriously considering only one lender, while 60% of borrowers who apply to 5+ lenders seriously considered only one lender. Share of population in each category listed above bars.

brokers/lenders, 48% of those who apply to four brokers/lenders, and 25% of those who apply to three brokers/lenders seriously considered *fewer* lenders than they ended up applying to. Overall, 16% of respondents with multiple applications report applying to more loans than they considered. This suggests that rejected applications and high rate draws, both of which should be counted as search and generate an inquiry, may be substantially under-represented in survey data which only solicit options that were seriously considered. While such survey data are invaluable for studying how decisions are made given a set of approved loans with a narrow band of drawn rates, they may be relatively uninformative about search in our setting with application rejections.

Nevertheless, a substantial share (28%) of respondents who apply to multiple lenders say they do so out of concern for qualifying for a loan. This provides suggestive evidence for the salience of our model's mechanism for borrowers. Furthermore, borrowers who report being concerned about qualifying for a loan realize statistically significantly higher interest rates on average than do borrowers who seriously consider multiple lenders/brokers for other reasons.

APPENDIX C: ADDITIONAL RESULTS

Figures A2–A4 present robustness of the key empirical facts of the paper, namely that realized interest and default rates increase in borrower search. Figures A2 and A3 plot the interest rates against search for a host of borrower subsets—by FICO, education, income, product type, loan purpose (i.e., refinance versus purchase), for the subset of loans that do/do not eventually default, and for the set of borrowers who do/do not obtain their



Panel A: By FICO Score

Panel B: By Application Date

FIGURE C1.—Relationship between search and mortgage application approval rates, conditional on observables by FICO bucket. *Notes*: Figure plots the relationship between application approval rate and the number of inquiries on an applicant's credit report in the application data. Panel A plots the relationship by applicant FICO score and Panel B plots the relationship by application date. A line of best fit, weighted by the number of applicants with s inquiries, is drawn as a visual aid.

mortgage from a broker. All panels except Panel A of Figure A2 plot the estimated regression coefficients from equation (10), estimated within subsets. Panel A of Figure A2 plots the unconditional relationship between search and rates by FICO bucket. Panel E of Figure A3 increases our set of controls to include every bucket for loan-level price adjustments (LLPA) provided by Fannie Mae.[3] In all cases, we find the positive or U-shaped relationship between search and interest rates.[4] Figure A4 re-estimates the relationship between search and default following equation (11) for these many borrower subsets and including the LLPA controls. Overall, the central facts of the paper appear robust to all manner of control variables and across nearly all subsets of borrowers. Table A1 reports summary statistics for our Loan and Application datasets across three different time periods, while Table A2 reports coefficients from regressions in which the dependent variable is the # of inquiries (column 1) on a realized loan (Panel A) or application (Panel B), or the probability of being in each quartile of inquiry counts (Columns 2 through 4).

Figure A5 plots moments from the estimated model and the data. Figure A6 reports maximum likelihood estimates of the model parameters, when we re-estimate the model within subsamples. Figure A7 shows the distribution of realized interest rates (Panel A) and search (Panel B), as well as the relationship between search and realized interest rates (Panel C), default probabilities (Panel D), application approval probabilities (Panel E), and the share of high type borrowers as a function of realized interest rates, both in our baseline model estimation and our counterfactual exercises.

Figure C1 plots the relationship between application and approval and inquiry counts in our applicant data. Panel A plots the relationship by FICO bucket, while Panel B plots the relationship across application date bins: before the house price peak of 2006q4, during the crisis of 2006q4–2009q4, and after 2009q4. Consistent with our model's predictions, we find a strong negative relationship between search and application approval in each subsample.

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

[4]The exception is amongst Adjustable Rate Mortgages (ARMs) for which the origination interest rate is a very noisy measure of the loan's cost.

APPENDIX D: COMPUTATIONAL APPENDIX

D.1. *Likelihood Construction*

In our model, an inquiry is a draw from the offered rate distribution. Let S_i denote a random variable equal to the number of inquiries on loan application i, and let A_{is} be an indicator for whether an application sent on the sth search was accepted. Define R_i to be the realized rate on mortgage i, and R_i^* to be borrower i's reservation rate. Let D_i be an indicator for whether borrower i defaults on the mortgage. We proceed using a maximum likelihood approach.

Consider the probability that a realized loan with s inquiries and origination interest rate r is observed. For the loan to have been realized on the sth inquiry, the borrower must have failed to originate a mortgage on their first $s-1$ inquiries, then observed a loan offered at rate r, applied for it, and had their application approved. To build the likelihood for such a borrower, suppose first that one could observe both the borrower's underlying type z and reservation rate r^*. The probability that the borrower originates a loan at a rate below r on their sth inquiry is the probability that the borrower (1) finds a draw less than r and applies for it on rate draw s, (2) has their application approved on the sth rate draw, and (3) did not originate a loan on the previous $s-1$ draws. Thus, the probability that the borrower originates a loan at a rate below r on their sth inquiry is

$$\Pr\{R \le r, S = s | z, r^*\} = \mathbf{1}\{r \le r^*\} \cdot p_z H(r)\big(1 - p_z H(r^*)\big)^{s-1}$$

for $\mathbf{1}\{X\}$ an indicator function equal to 1 if X is true. One may take the derivative of the above expression with respect to r to derive a likelihood of realizing a loan at rate r after s inquiries, conditional on a borrower's type and reservation rate:

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

for $h(r)$ the probability density function (pdf) of the offered rate distribution evaluated at r. We do not observe the borrower's reservation interest rate r^* or type z. Thus, to form a feasible likelihood, it is necessary to integrate over the borrowers' possible reservation rates and type. This yields the likelihood function for the joint distribution of origination rates and search:

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

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

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

Since the default indicator D_i is assumed to be independent from search and acceptance decisions conditional on borrower type, including this information into our likelihood function is straightforward. A borrower of type z, who has seen a share t/T of their loan term elapse 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 $(S_i, R_i, D_i|t, T)$ is

$$
\begin{aligned}
& l^L(R_i, S_i, D_i|t, T) \\
&= \lambda\big(D_i\big(1 - x_h^{t/T}\big) + (1 - D_i)x_h^{t/T}\big)p_h h(r)\int_r^\infty \big(1 - p_h H\big(r^*\big)\big)^{s-1}\,dF_h\big(r^*\big) \\
&\quad + \underbrace{(1-\lambda)}_{\Pr\{z=l\}}\underbrace{\big(D_i\big(1 - x_l^{t/T}\big) + (1 - D_i)x_l^{t/T}\big)}_{\Pr\{D_i|z=l;t,T\}} \\
&\quad \times \underbrace{p_l h(r)\int_r^\infty \big(1 - p_l H\big(r^*\big)\big)^{s-1}\,dF_l\big(r^*\big)}_{\Pr\{R_{is}=r, A_{is}=1, s-1 \text{ Failed Searches}|z=l\}}. \qquad \text{(D1)}
\end{aligned}
$$

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

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

It is straightforward to show, following a similar argument to above, that the numerator is

$$
\begin{aligned}
& \Pr\{s-1 \text{ failed inquiries} \cap \text{Applied}\} \\
&= \lambda\int H\big(r^*\big)\big(1 - p_h H\big(r^*\big)\big)^{s-1}\,dF_h\big(r^*\big) \\
&\quad + (1-\lambda)\int H\big(r^*\big)\big(1 - p_l H\big(r^*\big)\big)^{s-1}\,dF_l\big(r^*\big). \qquad \text{(D2)}
\end{aligned}
$$

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

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

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

$$\Pr\{s-1 \text{ failed inquiries}|\text{Applied}\} = \frac{\Pr\{s-1 \text{ failed inquiries} \cap \text{Applied}\}}{\sum_{1\leq s'\leq \tilde{S}} \Pr\{s'-1 \text{ failed inquiries} \cap \chi_{is'}=1\}}. \quad \text{(D3)}$$

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

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

Letting $\tilde{S}$ go to infinity and substituting back into (D3) yields the likelihood contribution of an application with s inquiries:

$$l^A(S_i = s) = \frac{\Pr\{s-1 \text{ failed inquiries} \cap \text{Applied}\}}{\lambda/p_h + (1-\lambda)/p_l}, \quad \text{(D4)}$$

where the numerator is defined as in equation (D2). Combining this with the likelihood of each realized loan from equation (D1) yields the likelihood for our full data.

To estimate our parameters, we maximize the log-likelihood for our sample of loans and applications. We assume that an approved loan application is reported in our loan-level data set with i.i.d. probability $q(X_i)$, where X_i are borrower characteristics. We thus allow there to be differences over time in the observables of borrowers in the market, but we consider q to be a nuisance parameter whose estimation is not of interest. Let the set of observations in the realized loan data set be given by $\mathscr{L}$, and the set of observations in the application data set be given by $\mathscr{A}$. We maximize the following log-likelihood with respect to a choice of a parameter vector θ:

$$L(\theta; q) = \sum_{i\in\mathscr{L}} \bigl[\log q(X_i) + \log l^L(R_i, D_i, S_i|\theta, t, T)\bigr] + \sum_{i\in\mathscr{A}} \bigl[\log\bigl(q(X_i)\bigr) + \log l^A(S_i|\theta)\bigr],$$

where $l^L(R_i, D_i, S_i|\theta, t, T)$ is given by equation (D1), and $l^A(S_i = s|\theta)$ is given by equation (D4). Since $q(X_i)$ is additively separable from θ, its value will not affect our optimal choice of $\hat{\theta}$. To uniquely identify the parameters, we impose that $p_h \geq p_l$.

To prepare the data for estimation, we residualize observed interest rates to reflect information that the lender can observe about the borrower without an in-depth screening. Following equation (10), we regress origination interest rates on the borrower's sex, race, age group, education, income group, and debt-to-income group, as well as origination year and property state fixed effects. As a result, our estimates should be interpreted as allowing lenders to price discriminate along easily observable characteristics. Second, we winsorize all applications with more than 11 inquiries, in order to match the maximum number of inquiries observed in the realized loan data set.

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 $H(r)$ and the reservation rate distributions for high and low types $F_h(r^*)$ and $F_l(r^*)$. To ease the estimation burden, we make two simplifying assumptions. First, we assume that high and low type borrowers draw their search costs from the same distribution $G(c)$. This assumption guarantees that the reservation rate distribution for each type is determined by the distribution of search costs and offered rates. To see this, recall that a type z borrower has the following relationship between their search cost c and reservation rate r^*:

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

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

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

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

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

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

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

for μ, σ the mean and standard deviation parameters of the underlying normal distribution. This assumption permits the analytical construction of the reservation rate distribution for high and low type individuals, and is motivated by the roughly normal distribution of residualized realized rates observed in Figure A1. Our final parameter vector is then $\theta \equiv \{p_h, p_l, x_h, x_l, \lambda, \beta_H, \beta_C\}$.

Suppressing the subscript H on the parameters of the normal for presentation, and letting $pdf_{\mathcal{N}(\mu,\sigma^2)}(x)$ and $cdf_{\mathcal{N}(\mu,\sigma^2)}(x)$ be the pdf and cdf of a normal distribution with mean μ and standard deviation σ evaluated at x, we have

$$\begin{aligned}\psi_z(r^*) &= p_z r^* H(r^*) - p_z \int_{-\infty}^{r^*} \frac{r}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(r-\mu)^2}{2\sigma^2}\right] dr \\ &= p_z r^* H(r^*) - p_z\left[\mu cdf_{\mathcal{N}(\mu,\sigma)}(r^*) - \sigma^2 pdf_{\mathcal{N}(\mu,\sigma)}(r^*)\right],\end{aligned}$$

where the second equality follows by integration by parts. The above expression may be numerically inverted in a computationally efficient way. Also the derivative of $\psi_z(r^*)$ is

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

which is computable given our approximation to $H(r)$. Thus, we may construct the distribution of reservation rates for a type z given our approximation of $G(c)$ and $H(r)$.

Finally, as a robustness exercise, we suppose that inquiries measure $s = s^* + \epsilon$, where s^* is the true number of searches and ϵ is i.i.d. measurement error distributed according to $\Upsilon(\epsilon)$. For estimation, we parameterize $\Upsilon(\epsilon)$ as a geometric distribution with success probability ν, allowing for a zero measurement error also with probability ν, conditional on the fact that $s^* \geq 1$. To incorporate this into our likelihood function, we integrate over this distribution:

$$\Pr\{S_i = s|z, r^*\} = \sum_{\epsilon=0}^{s-1} H(r^*)(1 - p_z H(r^*))^{s-\epsilon-1} \frac{(1-\nu)^\epsilon \nu}{[1-(1-\nu)^s]}.$$

For robustness, we re-estimate the model assuming that ν is equal to either 0.5 or 0.2.

D.2. *Estimating Supply Side Parameters*

As detailed in Section 2.3, we transform the interest rate setting problem into a discrete choice problem, in which lenders choose from a menu of K discrete potential rates to offer. This approach leads to the offered rate choice probabilities expressed in equation (4):

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

In equilibrium, this offered rate distribution must be consistent with the offered rate distribution $H(r)$ used to calculate the market shares expected from choosing rate r, as determined by (5). Furthermore, the maximum likelihood estimates of $H(r)$ must align with these choice probabilities. This suggests an approach to estimating the supply side parameters by minimizing the distance between our maximum likelihood estimates of $H(r)$ and the choice probabilities given by equation (4). Specifically, we choose the cost of making a loan m and variance of profit shocks σ_ξ in order to minimize the squared distance between the mean and variance of the maximum likelihood implied offered rate distribution and the logit-choice probability distribution.

We assume that borrowers default at a constant hazard, so the probability that a type z borrower with loan of term T survives through t periods is $x_z^{t/T}$ and lenders expect to reclaim a fraction $\tilde{x}_z = (x_z - 1)/\log(x_z)$ of a dollar loaned to a type z borrower.[5]

[5]Suppose a borrower originates a mortgage whose term is T, requiring N discrete payments of equal size. Letting $\Omega(t)$ be the survival function after a fraction t of the loan's life, we have that the expected repayment is $\sum_{1\leq n\leq N}\Omega(nT/N)/N$. Substituting in for $\Omega(t)$ using the proportional hazard assumption implies that the expected repayment is $\frac{1}{N}\frac{x_z^{\frac{1}{N}}(1-x_z)}{1-x_z^{\frac{1}{N}}}$. Taking the limit as N tends to infinity yields the result.

D.3. *Computing Counterfactual Offered Rate Distributions*

Since both the market share equations (5) and (6) and reservation rate distributions depend on the distribution of offered rates in the market, a lender's optimal offered rate choice $\hat{r}$ will depend on the choices of all other firms in the market $H(r)$. In equilibrium, the distribution of offered rates implied by the lenders' profit maximization problem $\hat{H}(\hat{r})$ must be the same as the distribution of rates $H(r)$ used to calculate a lender's market share functions. Thus, we need to solve a functional fixed point problem for $H(r)$.

Our approach proceeds in three steps. First, we guess a normally-distributed equilibrium offered rate distribution $H(r; \beta_H)$. Next, we use equation (4) to calculate an implied distribution of optimally-offered rates $\hat{H}(r; \beta_H)$. Finally, we minimize the distance between $H(r; \beta_H)$ and $\hat{H}(r; \beta_H)$ with respect to β_H. The problem may then be written as

$$\min_{\beta_H} \| H(r; \beta_H) - \hat{H}(r; \beta_H) \| \tag{D5}$$

for some appropriately chosen norm ||. We solve this problem using numerical gradient-descent optimization algorithms implemented with KNITRO, and match the mean and variance of the implied distributions to those of the guessed distribution.[6]

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

A second concern arises from numerical approximations. We approximate the equilibrium offered rate distribution with normal distributions, which are then fed into the market share equations in order to calculate logit choice probabilities for every feasible rate. The objective function in the minimization problem (D5) therefore compares a normal distribution with logit-implied choice probabilities. To evaluate the severity of this concern, we search for an equilibrium using the set of parameters estimated using our maximum likelihood routine. The mean and standard deviation of the MLE offered rate distribution are 0.142 and 0.547, respectively. By comparison, the "equilibrium distribution," obtained by running these parameters through the equilibrium search routine described above, has a mean and standard deviation of 0.206 and 0.723, respectively. Although imperfect, we consider this error to be relatively small. After simulating the demand side of the model, this leads to a gap in average rates paid of 2.9bp, and an increase in search of 0.13 inquiries per borrower.

APPENDIX E: ALTERNATIVE MODEL: SIMULTANEOUS SEARCH

This section considers an alternative model of search in which a borrower samples multiple rates simultaneously as in the classic Stigler (1961) model. We show that the main results of the paper survive this alternative search protocol.

As in the main text of the paper, we assume that borrowers differ in their creditworthiness and lenders have an informative screening process which allows them to approve

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

type z borrowers with probability p_z. Borrowers differ in their search costs c_i which are drawn i.i.d. from a distribution $G(c)$. This model differs from the main text in that search is not sequential. Instead, we assume that borrowers commit to simultaneously sample and apply to k rates with replacement, rather than searching rates sequentially. The outcome of a borrower's search process is therefore an i.i.d. sample of rates $(r_1, \dots, r_k)$ drawn from the offered rate distribution $H(r)$, along with a set of application acceptance decisions. Finally, let r take on values in the interval $r \in [\underline{r}, \bar{r}]$.

Without loss of generality, order the sample of rates such that $r_1 \leq r_2 \leq \dots \leq r_k$. The borrower takes the lowest rate at which their application is approved. Let $r^{(a)} \in \{r_1, r_2, \dots, r_k, \infty\}$ denote the lowest rate drawn from $H(r)$ that is also approved by the lender, where we allow $r^{(a)} = \infty$ to represent the case in which all of the borrower's applications are rejected. Given this notation, the probability that a type z borrower receives rate r_s is equal to the probability that they are rejected for all lower rates, but accepted for rate r_s. This may be expressed as

$$\Pr\{r^{(a)} = r_s | z\} = \begin{cases} p_z(1-p_z)^{s-1}, & \text{for } s = 1, \dots, k, \\ (1-p_z)^k, & \text{for } r_s = \infty. \end{cases}$$

Borrowers choose the number of searches k in order to minimize the cost of making the loan:

$$\min_{k \geq 1} \underbrace{\mathbb{E}[r^{(a)} | k, z, r^{(a)} \neq \infty]}_{\text{Payoff if originate loan}} \cdot \underbrace{(1-(1-p_z)^k)}_{\Pr\{r^{(a)} \neq \infty | z\}} + \underbrace{(1-p_z)^k v_z}_{\text{Outside Option}} + \underbrace{c(k-1)}_{\text{Search Cost}}, \tag{E1}$$

where v_z is an exogenous outside option satisfying $v_z \geq \bar{r}$ so borrowers accept the highest offered rate. When $p_z = 1$, this model simplifies to the classic model of Stigler (1961).

The payoff to search is the expectation of the minimum rate drawn, taken with respect to the distribution of $r^{(a)}$ defined above. One can show that this expected payoff is

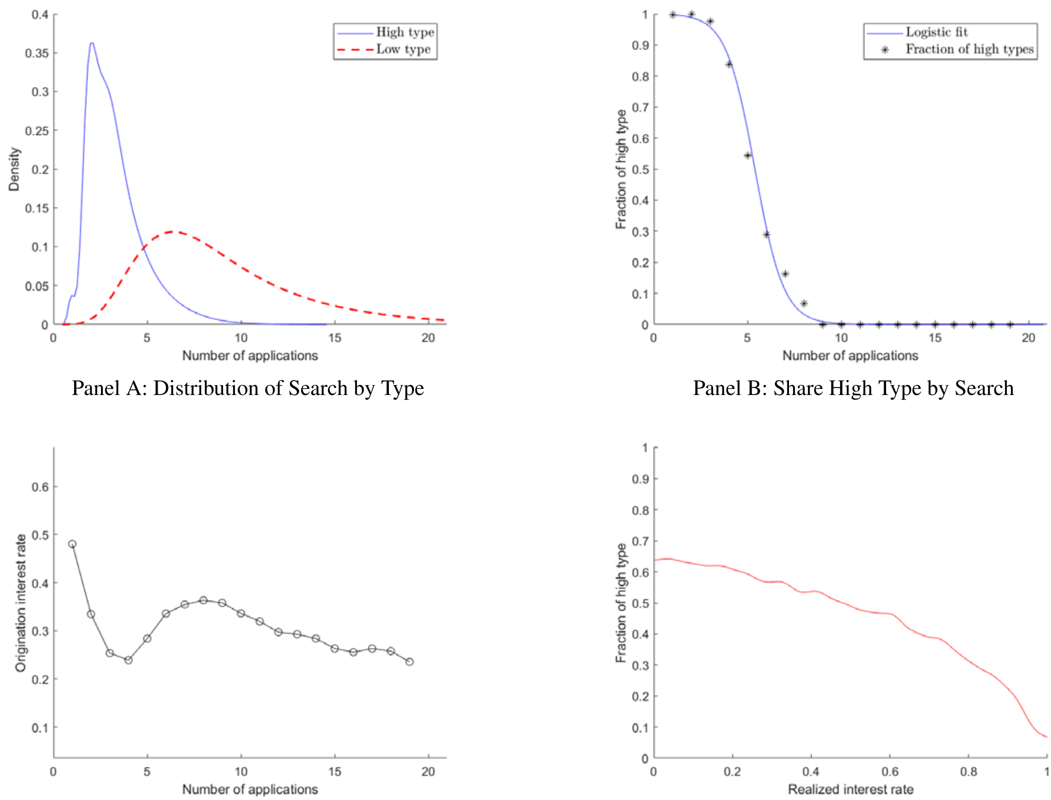
$$\mathbb{E}[r^{(a)} | k, z] = \sum_{j=1}^{k} \left(\int_{\underline{r}}^{\bar{r}} r h_j^k(r)\, dr \right) p_z (1-p_z)^{j-1} + (1-p_z)^k v_z, \tag{E2}$$

where $h_j^k(r)$ is the jth-order statistic of k draws of the offered rate distribution.

The borrowers' problem is not convex for all offered rate distributions. This is especially problematic, because the offered rate is endogenous, so a priori restrictions on functional forms are difficult to justify. This is partly why we limit attention to sequential search in the main text.

Nevertheless, the basic intuition of our model survives if borrowers search simultaneously. Consider the special case where rates are uniformly distributed between 0 and 1 so that $h_j^k(r)$ follows a Beta distribution. Under this assumption,the borrowers' problem is convex and admits a unique solution given a value of p_z, v_z, and c.

The intuition of why the relationship between interest rates and search is non-monotone mimics that of the sequential search model. In this model, the return to an additional search increases with p_z for two reasons. First, the probability that no application is approved falls with p_z. Second, the distribution of approved rates is more likely to contain at least one attractively low interest rate. As a result, borrowers with low approval probability tend to search more. However, because rates are only realized if approved, the probability that a low type borrower is able to realize a low rate mortgage is less than that of a high type borrower who searches the same amount. As a result, low type borrowers



Panel A: Distribution of Search by Type

Panel B: Share High Type by Search

Panel C: Realized Rates Versus Search

Panel D: Share High Type by Realized Rate

FIGURE E1.—Simultaneous search model simulation. *Notes*: Figure data simulated from a simultaneous search model with informative screening in which application approval parameters are set to $p_h = 0.999$ and $p_l = 0.2$, the share of high types is $\lambda = 0.5$, search costs are distributed according to a truncated normal distribution, and offered interest rates are distributed uniformly between 0 and 1. Panel A plots the distribution of search for high type (in the blue solid line) and low type (in the red dashed line) borrowers. Panel B plots the percent of borrowers that are high type at each realized level of search. Panel C displays the relationship between search and realized interest rates. Panel D plots the percent of borrowers that are high type at each realized interest rate.

sort to higher interest rates, just as in the sequential search model of the main text. The relationship between search and realized interest rates is therefore ambiguous.

Figure E1 plots output from a simulation of this model.[7] Panel A plots the distribution of search for high types (in solid blue) and low types (in dashed red). Search is defined as the number of rates a borrower samples k. Low type borrowers search much more than high type borrowers. Therefore, those who have many applications are much more likely to be low type borrowers: Panel B plots the share of borrowers who are high type against the number of searches chosen, which is tightly connected to the relationship between search and both default and application approval rates.

[7]We assume that search costs are distributed according to a normal distribution with mean 0.04, standard deviation 0.02, truncated to range between 0.01 and 0.08. High types are nearly always accepted, while low types are rejected at a high rate: $p_h = 0.999$; $p_l = 0.2$. High types constitute half of the population: $\lambda = 0.5$.

This sorting generates a non-monotone relationship between search and interest rates, shown in Panel C. Only high type borrowers choose a low level of search. Likewise, only low type borrowers choose a high level of search. Conditional on borrower type, those who search more tend to obtain better rates. Therefore, we see the usual downward sloping relationship between search and interest rates. However, for moderate levels of search, both high and low type borrowers search. The influx of low type borrowers raises the average rate paid as we increase search from low levels. Therefore, the simultaneous search model generates the same non-monotone relationship between search and realized interest rates that is observed in the sequential search model of the main text.

Finally, Panel D plots the share of high type borrowers against the realized interest rate in the market. The simultaneous search model generates the same adverse selection patterns found in the sequential search model. As interest rates rise, the borrower pool selects towards low types who are routinely rejected from low interest rates. This holds even though we assume that both high and low types have the same outside option: $v_h = v_l = 1$. The main predictions of our model are robust to a setting with simultaneous search.

REFERENCE

STIGLER, GEORGE J. (1961): "The Economics of Information," *Journal of Political Economy*, 69 (3). [21,22]

Co-editors Aviv Nevo and Guido W. Imbens handled this manuscript.

Manuscript received 3 June, 2020; final version accepted 10 March, 2024; available online 18 March, 2024.