Free Riding on the Search of Others: Information Externalities in the Mortgage Industry*

George Deltas† Zening Li‡

April 6, 2017

"BIGLY" PRELIMINARY.

PLEASE DO NOT CIRCULATE OR TWEET.

WE WILL UPDATE FOLLOWING THE CONFERENCE.

Abstract

Consumers often learn prices by engaging in costly search. Their search confers benefits to themselves, through lower prices. We show that it also confers positive externalities to other consumers in the same market. These can be either direct, by sharing the information from prior searches and thus improving the effectiveness of the search process, or indirect, by changing the equilibrium strategies of firms. Either type of externality allows consumers to reduce their own (costly) search activity: a consumer in a market populated with low search cost consumers searches less than an otherwise identical consumer in another market populated with high search cost consumers, while obtaining lower prices. We present evidence in support of the presence of both direct and indirect externalities in the U.S. mortgage issuance industry, though the evidence is stronger for the former than for the latter. Given that in the mortgage industry banks tend to charge the same mortgage rates within all markets in a state, indirect externalities operate through the composition of active firms in a market. This margin is typically ignored in the theoretical literature, as often is the possibility of direct information externalities.

JEL Classification: D43, D83, D12.
Keywords: Consumer Search, Information Spillovers.

*We would like thank Alexei Alexandrov, Gene Amromin, Dan McMillen, Leonard Nakamura, and seminar participants at the Society for Economic Measurement Conference and at the University of Illinois for useful comments.
†Department of Economics, University of Illinois, 1407 W. Gregory, Urbana, IL 61801, United States. (deltas@illinois.edu)
‡Department of Economics, University of Illinois, 1407 W. Gregory, Urbana, IL 61801, United States. (zli54@illinois.edu)

1 Introduction

Product prices are often not freely observable. As noted in a voluminous literature, starting with Stigler (1961), consumers may be aware only of the price distribution and must engage in costly information acquisition to obtain specific price quotes. The information acquisition often takes the form of search, where a consumer balances the cost of obtaining an additional quote, versus the benefit of discovering a lower price.[1] Consumers vary in search costs. Therefore, different consumers facing the same price distribution differ in their search activity. Because markets consist of different proportions of various consumer groups, the search activity in some markets may exceed that of other markets even if the price distribution in both markets were the same. The former set of markets could be referred to as having a high baseline search propensity, while the latter could be referred to as having a low baseline search propensity. This baseline search propensity features prominently in the comparative statics we undertake in this paper.

Suppliers also often vary in their costs, and thus differ in their optimal prices under a given set of demand conditions. For each supplier, the search activity of consumers and the posted prices of competing suppliers yield a demand curve. The supplier sets a price that is a function of his cost. In the market equilibrium, the price and search distributions are such that (a) no consumer wants to change their search activity, (b) no supplier in the market wants to change their price, (c) no supplier who is in the market wants to exit (or a supplier who is not in the market wants to enter).[2] Most theoretical work in search takes the set of firms as given, and thus ignores the last element of the market equilibrium. However, this element is often of empirical relevance, and we will return to it shortly.

When deciding on his or her search activity, a consumer only weighs the private benefits against the private costs. But search activity also generates externalities to other consumers, as has been recognized since the early contributions of Salop and Stiglitz (1977) and Varian (1980). One part of these externalities is direct. For example, consumers may share their experience with their social network, informing their friends about low price vendors to shop with and high price vendors that are best avoided. For multi-product vendors, this information may be valuable beyond the specific product that a consumer has bought: information that a consumer has found a good price for one product in a vendor may suggest to that consumer's

[1] Though our language may be suggestive of sequential search, we do not presume a specific search protocol. As Manning and Morgan (1982) point out, a fixed sample size (fixed n) search may often be superior, and an optimal search may contain features of both (see Morgan and Manning, 1985; Wilde and Schwartz, 1979, also provides some relevant discussion).

[2] Non-degenerate equilibrium price distributions do not require firm and/or consumer heterogeneity, as discussed here. Burdett and Judd (1993) derive conditions under which price dispersion arises even with homogeneous consumers and firms.

circle of friends that the vendor is a good prospect for other products. For example, if a borrower has secured a low interest rate car loan from a bank, some of his friends or relatives who are cognizant of this may consider that bank when searching for a mortgage. These direct externalities result in search that utilizes (imperfect) vendor-specific price information and is partially directed rather than random. It can thus be more efficient, as low price vendors may be "oversampled" relative to high price vendors.[3]

Another part of the search externalities may be indirect and operate through the market equilibrium. The prevalence of low search cost (high search intensity) consumers makes a market more competitive and prices generally lower. In fact, in the benchmark sequential model of Stahl (1989) the entire price distribution shifts to the left as the proportion of informed consumers increases; similar results are present in other models. These price effects can arise either from lower margins of firms active in the market, or by a change in the composition of firms towards lower cost vendors, or a combination of both. In turn, these effects can reduce the incentives of consumers to undertake search activity. Indeed, as Salop and Stiglitz (1977) have pointed out, high search cost consumers confer a negative externality to low search cost consumers, by making it possible for high priced firms to remain in the market: low search cost consumers would now have to search longer until they find a low price. Therefore, a consumer in a market with a high baseline search activity may search less than an otherwise identical consumer in a market with a low baseline search.

We ask whether there is any evidence for direct or indirect search externalities in the U.S. mortgage loan industry, and attempt to measure their importance. We investigate to what extent these externalities affect the search activity, the price distribution, and the set of firms active in each local mortgage market. We combine data on both search intensity and transaction prices by county for the 2013-2014 period; older data are used for the construction of instruments as appropriate. We find that externalities are present and non-negligible, though perhaps quantitatively not very important. Direct externalities appear to be stronger than the indirect ones, at least for this market. Because financial institutions tend to have uniform pricing within a state, indirect externalities manifest themselves only through the differential participation in high versus low priced banks in different geographical submarkets.

Because of the nature of the questions we ask and the type of data in our disposal, we adopt a reduced

[3]This is a topic that has been studied by the marketing literature, with Brown and Reingen (1987) among the early formal contributions. The idea of product information being exchanged "over the clothesline" and "across backyard fences" dates to Whyte (1954).

form approach.[4] We identify the search propensity associated with each borrower, property, and loan characteristic from the within-county differences in search rates. This part of the analysis utilizes data collected on the basis of the Home Mortgage Disclosure Act (HMDA), and contains a smaller set of variables than those available in the credit application. Importantly, however, it also contains data on some variables that may capture search activity differentials that cannot directly affect mortgage rates, such as race, ethnic background and gender. Some of the borrower characteristics have the expected association with search rates: all things equal, higher income individuals search somewhat less (possibly because of higher opportunity cost of time) and those who refinance search more (there is a less time pressure to secure a loan, and this is also a selected group of borrowers). There are no prior expectations for other characteristics, but we do estimate differential propensities based on borrower ethnic background and gender. Finally, the loan amount has a somewhat unexpected association with search intensity: those who apply for the smallest loans search the most, but beyond the first quartile the marginal effect of loan size is zero. A positive association might be expected, since larger loans imply larger gains from identifying a lower rate.[5] But a likely explanation is that the loan size itself may be correlated with borrower characteristics that are not included in the HMDA database, e.g., credit score, which may affect the level and dispersion of rates that these borrowers are quoted.

We use the composition of borrowers in each county and the coefficients from the regression described above to construct an index of baseline search activity (henceforth BSI). We then regress the county search rates, adjusted for the search propensity of its constituent borrowers, on the baseline search activity index (BSI) and other county characteristics and proxies that may be relevant for search activity, e.g., the concentration of the banking industry, per capita income, etc. In the absence of any search externalities, the coefficient of BSI would be zero, unless it were correlated with unobserved factors that are relevant for search activity. However, the coefficient on BSI is consistently negative and significant. In other words, an individual in a county populated by borrowers with high search propensity will be searching less himself/herself. The estimated parameter suggests a nearly complete crowd out, a result to which we return to below.

We next relate the baseline search activity to the price distribution faced by borrowers in each county.

[4] A companion paper, Deltas and Li (2017), examines a different set of questions using a structural framework.

[5] This reasoning is reminiscent of that in Sorensen (2000), who finds that price dispersion is lower for "maintenance" drugs, presumably because consumers have greater incentives to search for low prices for drugs they will be purchasing for many years.

To do so, we utilize a database that includes the interest rates and most other relevant mortgage application characteristics from all loans securitized by Fannie Mae, Freddie Mac, and Ginnie Mae. We adjust the mortgage rate for all application characteristics to construct a residual mortgage "price." Because an institution's mortgage pricing generally does not vary within a state, the only channel for across county differences in the offered prices comes from the seller composition.[6] We show that banks with high prices are more likely to operate in low BSI counties, relative to banks with low prices. Recall that prices cannot be tailored to local market conditions; the bank can only choose whether or not to operate there. As a result, counties with high values of BSI have marginally lower offered mortgage prices, after controlling for other relevant county characteristics. More importantly, we next compute separate distributions for mortgage price quotes (following mortgage applications) and also for the mortgage prices accepted by borrowers. The distribution of mortgage price quotes is noticeably more responsive to BSI than the average lender price in a county, and the distribution of accepted mortgage prices is even more responsive. These findings suggest that search in areas with high BSI is more "efficient," especially given the nearly full crowd-out effect reported in the preceding paragraph. Prospective borrowers in high BSI counties seem to obtain quotes from lower priced institutions; if they get multiple quotes, perhaps one from their bank (or the one the realtor recommends), any additional quote seems to be coming from relatively lower priced banks, possibly based on information from their social network.[7] Indeed, consistent with this explanation, the effect of BSI on the quote and transaction price distribution is concentrated around or below the median of those distributions.

Our study has a number of implications. The first implication is that search externalities lead to a suboptimally low level of consumer search in the mortgage market. Search is a "public good" that aids price discovery, increases market competitiveness, and pushes inefficient producers out of the market. Thus, policies that reduce the search cost to consumers of mortgage services would be welfare enhancing if those policies have some modest costs. Second, high search groups provide an implicit transfer to co-located low search groups, i.e., as noted above search externalities have a redistributive element. Lastly, the presence of

[6] This geographical price uniformity, at least within individual states, is a feature of the U.S. market and derives from the prohibitions of "red-lining." This prohibition is often based on the premise of non-discrimination, but can also be justified on the basis of correcting privately optimal but socially suboptimal lender behavior (see Lang and Nakamura, 1993). In other countries, mortgage rates may be negotiated at the individual level (see Allen, Clark and Houde 2016). In areas where there is no heterogeneity in risk or information for mortgage issuance, multi-branch banks may prefer to commit to uniform pricing (see Calem and Nakamura 1998). In this case, fear of litigation serves as a commitment devise that increases equilibrium profits of multi-branch banks.

[7] The bank from which a particular consumer gets a quote deterministically (i.e., with probability one) is "prominent" in the terminology of Armstrong, Vickers and Zhou (2009). However, in the mortgage market, the identify of this prominent bank differs from consumer to consumer, even within a specific county.

direct search externalities might pose a problem for simple structural search models. A low observed transaction price may due to a “tip” from friends/family and not the outcome of extensive search. Transaction prices are lower than implied by the number of searches, unless provision ii made in the model to allow for some searches to be partially informed.

This paper is related to the empirical literature on search, most of which uses structural empirical methodologies to estimate the market fundamentals and perform counterfactual simulations. One relevant to our study part of this literature focuses on the mortgage industry and investigates issues of price dispersion and search frictions. Gurun, Matvos, and Seru (2013) find large differences (2.8 percentage points between the 95th and 5th percentile lenders on average) in reset rates for adjustable-rate mortgage (ARM) loans charged by lenders within geographic regions after conditioning on borrower and loan characteristics and the initial interest rate. Alexandrov and Koulayev (2017) provide some direct evidence on the extent of mortgage price dispersion, obtained through the use of a unique proprietary data of mortgage rate sheets collected by the CFPB and covers the most important lenders. Even the most competitive segments of the market exhibit a dispersion of 0.5%, which is similar to the price dispersion we obtain in our study using a different dataset and methodology.[8] They then proceed to calculate the consumer gains if the fraction of consumers who search increased by a substantial (but plausible) amount. Allen, Clark, and Houde (2016) suggest that search frictions in the Canadian mortgage industry reduce consumer surplus by almost $12 per month per consumer, and that 28% of this reduction can be associated with discrimination, 22% with inefficient matching, and the remainder with search cost. In a similar vein, Woodward and Hall (2012) points out that confused consumers overpay their brokers’ services at least $1,000 by shopping from too few brokers. Perhaps due to its financial complexity, searching for a suitable mortgage does seem to be a daunting task for most consumers, a finding that is implicit in the findings of our work.

A second relevant to our study component of the empirical search literature estimates the direct effects from consumer search and contrasts them with the strategic effects of search, i.e., the effects that increased search has on the price distribution. This latter effect is what we refer to as indirect externalities. Recent work by Brown (2016) shows that for relatively small search levels, such as those for medical imaging procedures in New Hampshire, the direct gains from search dominate the strategic ones, but for high levels of search the reverse is true. Another recent contribution, by Salz (2015), finds evidence that in the New

[8] Price dispersion in what are homogeneous financial products is not limited to mortgages, as Hortacsu and Syverson (2003) point out in their study of mutual funds that track the S&P index.

York trade waste market, those who use intermediaries (brokers) to identify low price providers confer an externality to those who do not use intermediaries. This finding explicitly links the effects of increased search on the price distribution to a redistribution of surplus from the firms to the consumers who choose not to search.

2 DATA

The data we use come from two main sources, and are also complemented with information from other data series. The first source is the annual loan application register data provided by the Home Mortgage Disclosure Act (HMDA) of 1975, which requires many depository and non-depository lenders to collect and publicly disclose information about housing-related loans and applications. Whether a lender satisfies HMDA's coverage criteria depends on its size, the extent that its business is in an MSA, and whether it is in the business of residential mortgage lending.[9] In 2014, $7,062$ institutions reported data on nearly 10 million home mortgage applications, which is the vast majority of all applications made during this time.

Lenders report to HMDA the action taken following a loan application. Table 1 lists the number of observations corresponding to each of these actions for first-lien, $1-4$ family homes mortgage applications.[10] Of all the applications in 2014, nearly 61% resulted in an origination. The second largest action category is denial of the application by the lender, which is 17.71% of all applications reported. Denied applications are not considered as searches for the purposes of our study because they do not yield an option to borrow funds. Moreover, applicants who get turned down by a lender during their first mortgage application typically do not apply to another institution.[11] For the same reason, incomplete applications are also not considered valid searches. Applications approved but not accepted and applications withdrawn by the applicant sum up to more than 15% of the total applications. These two actions are considered as searches that borrowers made, and after comparison with their other options, chose to turn down.[12] HMDA also

[9]Specific regulatory disclosure requirements are available in the HMDA Guide on page 26. https://www.ffiec.gov/hmda/pdf/2013guide.pdf

[10]The current version of 2014 HMDA data that is downloadable from the CFPB website https://www.consumerfinance.gov/data-research/hmda/explore is slightly different from what we downloaded in 2016 in terms of transaction counts. This is because the 2014 data were updated to include approximately 174,000 transactions from Green Tree Servicing, LLC's HMDA submission, which were not incorporated into the public 2014 data until after the dataset was finalized.

[11]See Mondragon (2015). The reason for this behavior may be that the applicant infers from the denial that they are not creditworthy and will also be turned down by other institutions.

[12]The top 2 reasons for withdrawing an application is that the borrower obtained a better quote or more timely funding from other sources (http://linear-title.com/top-reasons-a-borrower-might-withdraw-their-mortgage-application/), while the fourth is that the process takes too long with that specific lender. Changes in the borrower's circumstances are another common driver for application

has selected data on pre-approval requests, which we use in robustness checks.[13]

HMDA also contains information on borrower characteristics, such as the applicant's race, gender, ethnicity, and income, and also information on loan characteristics, such as the loan amount and purpose, owner occupancy, loan type, property type, and lender ID. Nearly all of the applications are available at the census tract level, where the location refers to that of the property, but we use county level data in our analysis. The main reason is that the census tract is too narrow a geographical area for defining retail mortgage markets. Amel, Kennickell, and Moore (2008) find that the median household lived within four miles of its primary financial institution and 25% of households obtained mortgages from their primary financial institution. Moreover, more than 50% of households obtained mortgages from an institution less than 25 miles away. This indicates that the majority of borrowers search at least one local mortgage lender if not more. They might still use other resources to obtain additional information or price quotes, but it seems like they tend to originate their loan with a local lender. Borrower location will often not correspond to property location, which is the geographic information in our data, because many home purchases take place when people move across counties, and this might explain a portion of the mortgages obtained from banks that are not in close proximity with the borrower. Note that given the nature of the geographic identifier in HMDA, we define markets based on the location of properties, not the location of borrowers. Of course, since the properties in our sample are for most part owner occupied properties, following the purchase owners and properties will typically be in the same location; but this will often not be the case when the mortgage application is filed.

We show our data selection process in Table 2. Starting with 8,973,748 mortgage applications for first-lien, 1-4 family homes, we first drop applications that are not categorized as searches, i.e. applications that are denied by the financial institution, closed due to incompleteness, or are pre-approval requests. This leaves us with approximately 6.8 million searches in 2014. Since our analysis is at the county level, we then drop applications with county code missing. Eventually, we drop applications that are "not applicable" to race, gender, ethnicity, or owner-occupancy questions. This usually indicates that the applicant is not a natural person (for example, a corporation). Since the loan terms, approval criteria, loan characteristics, etc. for commercial mortgage loans are very different from residential mortgage loans, we think of them

withdrawals. We have not seen any mention that applications are withdrawn because of possible denial of credit.

[13]Doing so requires some adjustments to account for the incompleteness of the data. This is, in fact, the main reason why pre-approvals are not used for the main results.

as mortgage products in two separate markets. We drop these applications because our paper focuses on search behavior in the individual borrower residential mortgage loan market. After the above selection process, we have 6,700,772 applications remaining.

Of these applications, approximately a fifth did not result in originations. We consider those to be a measure of borrower search activity. It may appear striking that the number of searches per "purchase" is only 1.25, so some discussion is warranted. One reason for this low figure is that our definition of a search is rather stringent. For example, browsing the web to identify an institution's headline rates or to obtain a quote based on partial borrower information is not considered a search, largely because these rates may be weakly related to the actual rates a borrower would obtain if he/she filed an application. However, we recognize scouring the web for quotes has some value, as do pre-approval requests. Probably the most appropriate way to interpret our measure of search is that it is a proxy for overall search activity, since the number of these other "softer" quotes that borrowers obtain are likely strongly correlated with the mortgage applications they file to obtain hard quotes. Importantly, in our interpretation of our results, we will also consider our measure of search as a proxy for information acquisition from the borrowers' social network, which would provide both direct recommendations (or tips) or information about the lender's reputation.[14] The second reason why the average search figure is low is because prospective borrowers do not, in fact, engage in much search activity. According to the National Survey of Mortgage Borrowers (NSMB), almost half of consumers who take out a mortgage for home purchase fail to shop prior to application.[15] Moreover, for most borrowers, their mortgage shopping experience stops after their first application, as corroborated in the HMDA data. According to Lacko and Pappalardo (2007), in a survey conducted by the Federal Trade Commission, the typical borrower considered only two loans while shopping. These low search rates are despite the large dispersion in rates across financial institutions, even after accounting for loan size and mortgage type, as recently documented by the Consumer Financial Protection Bureau (CFPB)'s analysis of mortgage rate quotes.[16]

[14] In the 2015 study by the CFPB on *Consumers' Mortgage Shopping Experience*, available from http://www.consumerfinance.gov/reports/consumers-mortgage-shopping-experience/, these two factors were among the top 5 considerations for shopping from a particular lender.

[15] See *Consumers' Mortgage Shopping Experience.* for relevant figures.

[16] The CFPB notes that "shopping is important not only to help borrowers understand the different product features available, such as adjustable-rate versus fixed-rate, but also the price at which those products are offered." Recognizing the potential benefits of effective shopping, the CFPB is improving mortgage disclosures under the Truth in Lending Act and the Real Estate Settlement Procedures Act. In October 2015, the "Know Before You Owe" mortgage disclosure rule replaces four disclosure forms with two new ones, the Loan Estimate and the Closing Disclosure. The new forms are designed to help consumers understand their loan options, shop for the mortgage that's best for them, and avoid costly surprises at the closing table. To further encourage mortgage shopping, the CFPB has also launched various tools and resources that help consumers make more informed decisions during the mortgage

We assume borrowers who share the same observable characteristics across locations have similar search propensities, even though they will have different equilibrium search rates. That is, if these borrowers were placed in identical environments, their search activity would be similar; actual search activity will differ only because of market characteristics, e.g., the distribution of mortgage rates. Search "primitives" such as search costs, prior information about the mortgage market, and gains from incremental search depend similarly on observable borrower and property characteristics (and associated proxies) across markets. We first estimate how different characteristics affect search propensity and then use the composition of these characteristics to construct the baseline search intensity for markets.

In what follows, we assume that the bulk of applicants primarily search (for both properties and mortgages) within a county, report the same characteristics in all applications, consider properties and loans of a given type, and originate the loan within the calendar year if they were approved. Therefore, all applications within a county that share the same borrower, loan, and property characteristics will be filed by individuals by the same search intensity: their search propensity will be the same, given the assumption that it is a function of borrower, property, and loan characteristics, which results in the same equilibrium search intensity since these individuals are active in the same mortgage market.[17] The expected search activity in a particular market of an individual with a given set of values for these characteristics equals the total number of applications filed in that market by all individuals with this set of characteristics divided by the number of loan originations taken out by these individuals.

The list of characteristics that we distinguish is dictated by the availability of data in the HMDA database. We thus group applications in a county that have the same combination of values for race, gender, ethnicity, loan purpose, owner occupancy, loan type, loan amount level, and income level (the last two are continuous variables and are discretized as noted below). Race includes White, Black or African American, Asian, and others.[18] Gender includes male, female, and not provided. Ethnicity includes hispanic or latino, not hispanic or latino, and not provided. Loan purpose includes home purchase, home improvement, and refinancing. Owner occupancy includes owner-occupied as a principal dwelling and not owner-occupied as a

searching process. See http://www.consumerfinance.gov/owning-a-home/ for more details.

[17] We recognize that some borrowers will obtain financing from an institution that is not located in the same county as the purchased property. Noting that an institution that is present and active in a county will typically approve multiple mortgage applications in a single year, we do not include in some of our analysis institutions with a single mortgage approval.

[18] "Others" includes individuals for whom the field is not provided and those who belong to the two demographic groups of Native American and Alaska Native, Native Hawaiian and Other Pacific Islander. These two demographic groups have too small a number of observations to include separately in the estimation, and they have a high correlation with the not provided group (0.82 and 0.77, respectively).

principal dwelling.[19] Loan type includes Conventional (any loan other than FHA, VA, FSA, or RHS loans), FHA (Federal Housing Administration)-insured, VA (Veterans Administration)-guaranteed, and FSA/RHS (Farm Service Agency or Rural Housing Service). For loan amount level, the four types we defined are the four quartiles within each state. Income level types are defined in the same way, with an additional type containing those that did not report their income (around 5% of total observations in 2014). Table 3 reports the composition of individual/property/loan characteristics among the applications. Some combinations of group characteristics have zero observations for many counties. Year 2014 has 7,806 groups and 940,436 county-group level observations with the median number of groups per county being 162. On average, there are 5.71 originations and 7.12 searches for every county-group combination in our sample.

A special discussion of the "not provided" category may be useful. For some variables, mortgage applicants need not provide a response. By far, the highest percentage of non-response is for the race and ethnicity variables, where it is approximately a tenth of the sample. We suspect that a substantial number of individuals do not wish to classify themselves in one of the established categories (as an indication 9.3% of individuals in the 2010 US Census described themselves as being part of two races or of "some other race."). Others may prefer not to report a race on principle. For the purposes of our analysis, the group of individuals who do not report this information is treated as a distinct socio-economic group. More surprising is the fact that six percent of the applications do not list a gender. One can easily speculate on some possible reasons, but the "non-reported" category is also treated as a separate group in our analysis. Treating these individuals as distinct groups is reasonable if their behavior systematically differs from that of other groups, given that they are part of the market and their search activity has implications for the response of firms and for the search activity of other consumers. However, we have also re-estimated all specifications after dropping these individuals from our sample, and obtained similar conclusions.

We constructed the Herfindahl-Hirschman Index (HHI) using the lenders' mortgage origination share in each county. This index is used as an indicator of each market's concentration level. To address possible endogeneity concerns with lender concentration at the county level, we also constructed two instruments for HHI. The first is the HHI calculated from 2007 HMDA data. The second instrument is the increase in HHI between 2007 and 2014 that is attributable to the merger and acquisition activity of banks and bank holding companies, using 2007 market share data.[20] This activity is unlikely to be correlated with county-

[19]Second homes, vacation homes, and rental properties are classified as not owner-occupied as a principal dwelling.

[20]More precisely, we compute the HHI using 2007 market share data after combining the market share of all banks and bank holding

specific economic considerations, given that most banks operate over multiple counties. The M&A records are provided by the Federal Reserve Bank of Chicago and contain information that can be used to identify all bank and bank holding company acquisitions and mergers that have occurred since 1976.[21] Summary statistics on HHI and the associated instruments are reported in Table 4.

Our second major dataset contains all the fixed-rate conventional loans originated in the 50 states, Washington D.C., and Puerto Rico in 2014 (approximately 4.3 million individual loans) that are securitized by Fannie Mae, Freddie Mac, and Ginnie Mae.[22] The GSE securitized share of first lien origination volume was 52% and FHA/VA originations securitized by Ginnie Mae was 21%. Hence, in total our dataset accounted for 73% of the first lien origination volume in 2014. We dropped observations that have one or more than one of the following key variables missing: loan rate, credit score, LTV (loan-to-value ratio), DTI (debt-to-income ratio), loan purpose, loan amount, loan term, third party origination flag, number of borrowers, number of units, origination month, state, seller, and securitizer. The dataset also includes information on the occupation status and property type for GSE securitized loans and a first time buyer flag for 95% of the data. However, we do not have information on the points and fees borrowers pay. See Tables 5-6 for a summary of these loan characteristics.

We also obtain county level demographic characteristics from the American Community Survey's 2014 5-year estimates. It contains information on population, education attainment, household income, per capita income, worker population, employment rate, labor force, occupied housing units, rent units, median gross rent, and median housing value. These variables are used as controls for county characteristics in our analysis, and key summary statistics are reported in Table 7.[23]

---

companies that have merged between 2007 and 2014. We then take the difference between that "counterfactual" HHI for 2007 and the actual HHI for 2007. This difference is used in our analysis as an instrument for the HHI of 2014. A similar instrument is used by Scharfstein and Sundaram (2014).

[21]The data files were obtained from https://www.chicagofed.org/banking/financial-institution-reports/merger-data. Updated versions of the data will be available from the National Information Center Bulk Data Download page.

[22]All HMDA applications in 2014 come from the same locations: the 50 states, Washington D.C., and Puerto Rico.

[23]The Federal Deposit Insurance Corporation (FDIC) website has a listing of branch office locations and their annually reported deposits as of June 30, 2016. (Data is available back to 1994.) The listings provide branch office data by state, county, city and institution, downloadable at: https://www5.fdic.gov/sod/dynaDownload.asp?barItem=6. This database, which is not used in this version of the manuscript, may allow us to construct an alternative non-binary measure of bank presence in a county that does not directly depend application volume.

3 SEARCH ACTIVITY AT THE DEMOGRAPHIC GROUP & MARKET LEVELS

Our first task is to pin down the differential search rates by borrower-type, where borrower-type consists of the set of borrower, property, and loan characteristics reported in the HMDA database. Recall that we do not observe individual borrower identifiers, and thus we do not know how many mortgage applications were initiated by each borrower of a given borrower-type. What we do observe is how many mortgage applications and mortgage originations were performed by all borrowers of a given borrower-type for properties in a given county. We know, then, the number of borrowers of borrower-type j in county/market m, which is equal to the number originations $Orig_{j,m}$ attributable to borrowers of that type in county m. We also know the total number of searches performed by all borrowers of each borrowing type in a market, which is equal to the number of applications approved by the lender or withdrawn by the borrower, $Apps_{j,m}$. From these, we compute the average number of "searches" by members of borrower-type j in market m, $Apps_{j,m}/Orig_{j,m}$. As noted earlier, we consider this ratio to be a proxy for less formal rate queries and other information acquisition efforts.

This ratio varies across markets on the basis of borrower-type and market characteristics. We postulate that the effects of these characteristics are additively or multiplicatively separable, and that unobserved factors that affect the search ratio are not systematically related to the observed ones. Moreover, the behavior of individuals of the same borrower-type is assumed to be the same across counties that are otherwise identical. This last condition is clearly a substantial departure from reality, but there must be some large commonalities of behavior given the statistically significant findings.[24] These assumptions allow us to identify the differential search intensity associated with each borrower-type. In our framework, where the market characteristics are captured by a set of exhaustive dummies, borrower-type search propensity is identified from within market differences in per borrower search activity of each type.

Fixing some notation is useful in formalizing the analysis. Suppose we have K discrete borrower-type characteristics $T_1, T_2, \ldots, T_K$ and we denote the different values of T_k by $T_{k,1}, T_{k,2}, \ldots, T_{k,\kappa}, \ldots, T_{k,\kappa_k}$. Borrower-types, which to economize on words we will often simply refer to as "groups," are defined by elements of the form $\{T_{1,l_1}, T_{2,l_2}, \cdots, T_{K,l_K}\}$, where $l_k \in \{1, 2, \ldots, \kappa_k\}$ and where all possible combinations of $l_1 \times l_2 \times \cdots \times l_K$ are used. For example, a group would be defined with a specific combination of ethnic

[24] In fact, if each borrower-type is a mixture of underlying sub-groups with a composition that varies across markets, this would lead to an attenuation of our results.

and socioeconomic (categorical) characteristics, applying for a mortgage of specific maturity and type, on a property within a given price range and characteristics. Many of these combinations, however, contain no individuals for at least some counties. Denote T_{k_j} as the value of characteristic k for group j and $w_{j,m}$ as the proportion of group j in market m.[25] Out of the 6,700,772 individual level applications, there were 940,436 unique group-market observations. There are some group-markets with only searches but no loan originations, likely because the borrower purchased a home in a different calendar year, or in a different county, or his purchase plans otherwise changed. This decreased 940,436 group-market level observations to 812,446.

These observations form our dataset for the first step of our analysis, which explains the variation in $Apps_{j,m}/Orig_{j,m}$ as a function of group and market characteristics. The independent variables are all binary indicator variables. Those for group characteristics take the value of one if that group has a particular attribute and the value of zero otherwise. Market characteristics, which are not of direct interest in this regression, are treated in the most flexible way and consist of an exhaustive set of county dummy variables. These dummies provide the search activity of the omitted group in each county, but the additively scale the search activity of all groups in a county. We will refer to them as the market effects, or the adjusted-for-borrower-composition market level search activity. The regression we estimate is given by

$$\frac{Apps_{j,m}}{Orig_{j,m}} = \alpha_m + \sum_{k=1}^{K} \sum_{\kappa=2}^{\kappa_k} \beta_{k,\kappa} \mathbf{1}_{\{\kappa\}}(T_{k_j}) + \epsilon_{j,m} \tag{1}$$

where $\mathbf{1}_{\{\kappa\}}(T_{k_j}) = 1$ if $T_{k_j} = T_{k,\kappa}$, and 0 otherwise. The double sum consists of a series of dummies for each of the borrower-type characteristics, where one dummy per characteristic is dropped. Note that the independent variables take the exact same value for all the individual members of each group. Therefore, estimating this regression via weighted least squares, with weights equal to the number of individuals in a group, produces identical estimates to those we would have obtained had the individual-level data been available (and we had used OLS). A more efficient estimation approach is to estimate equation (1) via GLS, to account for the fact that idiosyncratic variability in search may systematically differ across group characteristics (including the size of the group).[26] GLS weights are almost linear in the number of individuals per group, so in this regard they do not depart much from analytic weights. But they do

[25]These proportions are calculated using originated loans.

[26]GLS is implemented via iteratively re-weighted least squares as follows: 1. Estimate the unweighted linear regression and obtain the residuals. 2. Regress the residuals on the number of originations at the group-market level and on all controls. 3. Obtain the predicted value of residuals. 4. Rerun the original regression with weights proportional to the reciprocal of the squared predicted residuals. Repeat step 2 and step 3 until the estimated coefficients converge. In this study, we did 5 iterations.

down-weight observations with attributes associated with high variance.

Because the group characteristics may affect search activity super-additively (possibly multiplicatively), we have also estimated the following log-linear specification of the above regression.

$$log\left(\frac{Apps_{j,m}}{Orig_{j,m}}\right) = \alpha_m + \sum_{k=1}^{K}\sum_{\kappa=2}^{\kappa_k}\beta_{k,\kappa}\mathbf{1}_{\{\kappa\}}(T_{k_j}) + \epsilon_{j,m} \tag{2}$$

Even though the results of these regressions are used primarily as an input to further analysis, we report them in Table 8. We observe some common features when comparing the linear and log specifications. For example, female borrowers search less than male borrowers, and both search less than those who do not report their gender. Hispanic borrowers search less than non-hispanic borrowers. Borrowers refinancing their mortgage search more than borrowers purchasing a new property, not surprising given that they are under less time pressure. If a mortgage loan is VA-guaranteed, it's searched less when compared to other loan types. Also, borrowers taking out mortgages with smaller loan amounts search more and borrowers with higher income tend to search less. There are also a very few discrepancies between these two models. For example, African Americans search the most across race groups in the linear model, while Caucasians search the most in the log model.

The coefficients of relative search activity associated with a particular group characteristic are not of ultimate interest. What is of ultimate interest is whether the activity level in a market, adjusting for its group composition, is systematically related to that composition in a way that suggests spillovers or free-riding. For example, suppose non-hispanic borrowers search more than hispanic borrowers in the same market, do borrowers in markets containing a high proportion of non-hispanics search less than identical borrowers in markets containing a low proportion of non-hispanics. This analysis could be performed by a "second-stage" regression, where the dependent variable are the estimates of the county dummies, α_m, from equations (1) or (2) and explanatory variables are the proportion of borrowers with each of the characteristics on the right hand side of those equations, plus any other market-level characteristics that can impact search activity. We would then compare the parameter estimates in "first-stage" equations (1) and (2) with those of this "second-stage" regression. We would expect that if a value of an attribute is associated with reduced search in the first-stage equations, the corresponding population weight coefficient in the "second-stage" would be positive.

However, this approach is fraught with two difficulties. First, there's too many characteristic coefficients to compare and the comparisons are not straightforward, e.g., a simple comparison of signs will not work because it is not invariant to the identity of the excluded category for each attribute. Second, some demographics may be proxies for other unobserved factors that affect search. Using a summary measure that combines all the estimates of the equations (1) and (2) eliminates the first difficulty. It also reduces the second, since it is unlikely that biases arising from the proxy effect of a demographic characteristic all point in the same direction. In particular, we use the group characteristics coefficients $\beta_{k,\kappa}$ and group market weights $w_{j,m}$ to construct every market's baseline search intensity BSI_m as:

$$BSI_m = \sum_{j=1}^{J} w_{j,m} \sum_{k=1}^{K} \sum_{\kappa=2}^{\kappa_k} \beta_{k,\kappa} \mathbf{1}_{\{\kappa\}}(T_{k_j}) \tag{3}$$

This index takes higher values for markets where borrowers have attribute values that are associated with higher market-adjusted search activity relative to borrowers with other attribute values.[27] Changing the identity of the omitted categories in equations (1) and (2) affects this index by a constant, i.e., it does not have an impact on the difference of this index across markets. When there are no interactions between attributes, as is the case in the analysis we report here, this index is identical to that obtained by multiplying the fraction of the borrowers that have a specific attribute value times the coefficient for that attribute value and summing over all attributes.

We then estimate the equation

$$\hat{\alpha}_m = a + bBSI_m + cX_m + u_m \tag{4}$$

where $\hat{\alpha}_m$ are the estimated market effects and X_m are other market characteristics that may associated with differential search levels. The effect of these characteristics may not be causal; rather, they may be proxies for causal factors. For example, causal factors that may affect search could be the density of bank branches, the traffic conditions that permit visiting those branches, or the competitiveness of the local banking services. Though we do not have data on many such factors, they are likely to be related to key economic and demographic characteristics in the market. The set of characteristics in X_m includes HHI, population, per capita income, the number of owner occupied units, the number of rent units, median rent, median housing value, percentage with a Bachelor's degree, percentage of minorities, percentage of workers, percentage of employment, and percentage of the labor force. As we mentioned in section 2, the current value of HHI is

[27] The search levels of some groups may be more responsive to the aggregate market search propensity, i.e., the assumption that county-level search activity is additively or multiplicatively separable from the composition of borrowers may fail. For example, some groups may be more sensitive to changes in the price distribution. This would create a potential bias in the coefficients of equation 1 and 2. But the value of the BSI_m may be less biased if group-specific biases cancel out.

instrumented with the 2007 value and the change in HHI driven by banking mergers and acquisitions since 2007. Observe that the dependent variable in equation (4) is an estimated parameter and thus contains sampling error. The standard error of the parameter, $\sigma_{\hat{\alpha}_m}$, is a measure of the sampling variability. Therefore, we estimate this regression using GLS, implemented via iteratively re-weighted least squares, where the error variance associated with an observation is a function of $\sigma_{\hat{\alpha}_m}$ and all the independent variables.

The results are shown in Table 9. We focus on the coefficient for BSI, which we interpret as reflecting the degree to which changes in the search propensity of the borrowers active in a market crowds out realized search activity. If there was no such crowd out, the estimate of b would have been zero: the search activity of borrowers of different characteristics would depend on market characteristics but not on the composition of borrowers in a market. A value of b in the $(0, -1)$ range implies that changing the composition of borrowers so that, holding everything else fixed, search activity increases by one unit, would in fact lead to a partial compensatory reduction in search activity by all borrowers in a market: overall search activity would increase, but not one-for-one. Finally, if $b < -1$, then overall search level in that market decreases. In our results, $b = -0.87$ for the linear model, suggesting that there's slightly more applications in markets with high search intensity individuals. For the log model, $b = -0.97$, suggesting that crowd out is about one-for-one and applications do not vary with the proportion of high search borrowers present in a market. With regards to other controls, a higher HHI, i.e. a more concentrated market, is related with less search. Search decreases when per capita income, employment percentage, and labor force percentage increase, possibly due to larger search costs for those employed with higher income. The number of owner occupied units, median rent, and median housing value can be interpreted as indicators of how active the local housing market is. We observe the amount of search for mortgages increasing with these three variables.

Before turning our attention to the link between search propensity and prices, it is worthwhile reemphasizing that the analysis of this section looks at a narrow definition of actual searches: filed applications. These could be a proxy for less involved search activity, that does not yield a binding mortgage offer by a bank. In that case, the estimate of b would be interpreted in the same way, if formal applications are more or less linearly related with informal inquiries. Word-of-mouth tips, however, may also serve the same role as "searches." If high search propensity individuals are also prone to asking individuals from their social circle for mortgage related information, this can substitute for formal searching as well. In that case, even if the observed coefficient for b is smaller than -1, the total amount of search activity in a market may increase with BSI. Search activity, including information exchange among borrowers' social circles, would leave a "signature" in the transaction prices. We turn to this next.

4 Search Propensity and Quote vs Transaction Prices

Search activity and information spillovers between consumers influence the transaction prices, holding the distribution of posted prices fixed. We proceed to measure the extent to which this is the case in the mortgage market, and draw inferences on the nature of search externalities. A major obstacle to doing so is that the HMDA database does not contain the rates offered by the financial institution. These must be obtained from other sources and combined with our HMDA sample. Unfortunately, there exists no publicly available mortgage rate data.[28] Therefore, we must "back-out" prices from mortgage transactions securitized by Fannie Mae, Freddie Mac, and Ginnie Mae.

Our starting observation is that in the United States, mortgage pricing is typically uniform within a state, because financial institutions fear possible exposure to allegations of "redlining." This observation has the following three implications. First, in "backing-out" firm pricing from transaction data, we can pool together all transactions involving a bank to the state level. Second, the search intensity of borrowers in a market has an attenuated effect on the prices a lender charges. These prices would reflect the competitive conditions in all the markets that a lender operates in and for many lenders, each individual county would be a small component of their total market. Therefore, third, search activity in a market will affect the distribution of prices available to borrowers in that market primarily through its impact on the presence of financial institutions in those markets, i.e., through the entry decision rather than through the pricing decision. Search activity will affect the quotes a borrower actually receives through the sampling probability of each lender. It will also affect the probability that a borrower accepts any given offered rate, i.e., it will affect the distribution of transaction prices given the distribution of price offers. In this section, we focus on the computation and comparison of these price distributions, while in the next section we look into the locational decisions of financial institutions.

In using the mortgage transaction data to back-out prices, we note that financial institutions offer interest rates based on information that is available on the mortgage application. All key items in that mortgage

[28] Alexandrov and Koulayev (2017) use a proprietary dataset collected by the CFPB.

application are available to us. However, banks typically offer to each borrower the opportunity to trade-off a lower interest rate with a higher upfront payment, known in the industry as points. The points chosen by each borrower are not available to us. Thus, a low observed interest rate in our transaction data may reflect a high points payment, and vice-versa. If borrowers with the same observable characteristics systematically chose different points depending on the financial institution that they transact with, this would render it impossible to ascertain which institutions are more expensive than the others. In what follows, we assume that the typical choice of points does not vary across institutions, conditional on borrower-type, and that the trade-off between points and interest rates is the same across institutions, i.e., institutions may vary in the rates they charge, but not in how rate discounts relate to points paid.

If these conditions are approximately met, then we compute the adjusted-price of an institution, which we will often simply refer to as price, from a regression of the transaction interest rate for a loan in the Fannie Mae, Freddie Mac, and Ginnie Mae database on borrower/loan characteristics and lender-state fixed effects. Formally, we estimate the equation

$$P_{i,l,s} = \mu_{l,s} + \zeta Z_i + e_{i,l,s} \tag{5}$$

where $P_{i,l,s}$ is the rate paid by borrower i to lender l in state s, $\mu_{l,s}$ are lender-state fixed effects, and Z_i is the full set of rate-relevant characteristics, including credit score, loan-to-value ratio, debt-to-income ratio, and others. The fixed effects capture the pricing of each lender in a state, after controlling for all other observable characteristics that might affect loan rates. The specification of equation 5 obtains an average measure of priciness of a financial institution for all borrower-types within each state. It is possible, in fact likely, that some institutions offer competitive rates for some types of borrowers (say those with high loan to value ratios) while other institutions offer competitive rates for other types of borrowers. This suggests that the bank dummies could also be interacted with some key borrower characteristics. However, we believe that the simple bank fixed effects provide an adequate measure of price dispersion in a market, even though the heterogeneity of pricing would be relevant for interpreting our results.

The results of the price regression are reported in Table 10. Except for the bank fixed effects, the regression coefficients are not further utilized in this paper. However, it is worth pointing out that the estimates are as expected, which is somewhat reassuring that the use of points is not very strongly correlated with borrower characteristics (and hopefully, then, not strongly correlated with the financial institution). Associated with a lower rate are, a better credit score, a larger loan, a mortgage associated with a home purchase, especially for a single-unit purchase. Associated with higher rates are high Loan-to-Value ratios and Debt-to-Income ratios, longer terms, and a retail transaction.

The estimates of the bank fixed effects are used to construct the price distribution in each county as follows. We first match the lenders in the price database with HMDA's respondent ID. We were able to match the lenders for 97% of our Fannie, Freddie, and Ginnie loans. This matching is crucial because county-level transaction volume information is only available in HMDA. We next compute, based on each lender's fixed effect, the lender's rate for a "typical" borrower, fixing the borrower characteristics to average national values. We refer to this rate, $p_{l,s} = \mu_{l,s} + \zeta \bar{Z}$, as the lender's adjusted price, or simply price. We then use the institutional fact that lender mortgage pricing is uniform within states to obtain the price distribution in each county. We first construct the raw distribution of (adjusted) price in market m by computing the fraction of lender's active in a market with $p_{l,s}$ below any value x. In calculating this raw distribution, each lender with at least one mortgage approval in a county gets the same weight. If prospective borrowers were equally likely to obtain a quote from any of the lenders operating in a county, then this distribution would be the relevant price distribution they would face. However, quite clearly some institutions are more likely to receive rate inquiries than others, if only because these institutions differ in size. A better approximation for the price distribution that consumers are facing could be obtained by weighting each banks price by its number of branches in the county (and making some similar provision for lenders operating via the Internet). We have not yet obtained branch data to compute this distributions, so for now we only discuss (later in this section) results based on the equally weighted prices.

We have also computed two other weighted price distributions. The first is based on the number of mortgage applications filed to lenders (and not rejected by them), which we refer to as the offer or quote price distribution. Let the number of applications to lender l in market m be $Apps_{l,m}$, and a lender's market share of applications by $s^A_{l,m}$. Then, the quote distribution, $F^A_m(x)$, is obtained by computing the fraction of market share weighted lenders with $p_{l,s}$ below any value x. The second weighted price distribution is obtained by using originations, $Orig_{l,m}$, to compute a lender's share of originations in market m, $s^O_{l,m}$. This is used in the same manner as $s^A_{l,m}$ to obtain the transaction price distribution, $F^O_m(x)$.

Table 11 provides summary statistics for these two price distributions. We highlight two statistics. First, the quote mean is higher than the transaction mean, but only by a small amount. High priced lenders should have a smaller shares in originations compared to their shares of quotes, since borrowers would naturally choose the cheaper provider if they obtain more than one quote. Therefore, a transaction weighted average price should be lower than the quote weighted average price. A small difference between the two would be expected since 80 percent of quotes end in originations. However, the difference is too small to explain by the conversion percentage. A possible explanation is that, as mentioned earlier, lender pricing does not respond to borrower characteristics in the same way. As a result, some borrowers may find one set of lenders to price lower for them, while other borrowers may find that another set of lenders offers the better prices. Therefore, search may lead borrowers to turn down some lenders for others, but the flows may partially cancel each other out. In other words, even though one lender may be generally more expensive than another lender, it does not necessarily follow that every borrower who gets a quote from both lenders will chose the later over the former.

The second statistic that we want to highlight is that the difference between the bottom and top decile of the quote distribution is approximately 0.3% (and similarly for the transaction distribution). In other words, there is substantial disparity in the average price of banks operating in a market. This figure compares with the 0.5% dispersion between the highest and lowest quote reported in Alexandrov and Koulayev (2017); if using difference between the top and bottom deciles, the corresponding difference would have been 0.4%, similar to our figure, despite the fact that we arrived at it through different methodologies: We used the full set of banks, backed-out their pricing from transaction data, and weighted them by market share, but we did not have the rate sheets in our disposal and were not able to account for points paid. This consistency is certainly reassuring.

We now examine how these price distributions vary across markets, and most important whether they are a function of a market's baseline search propensity measure that we constructed. Since lender mortgage pricing is uniform within states, any difference in the price distributions within states will be driven by the composition of lenders active in different markets and differences in their shares across markets. Holding the composition of active lenders the same, a market will have lower expected price quotes if borrowers in that market are more likely to file applications with lower price lenders. For each borrower, the transaction price equals the lowest of the quotes received. Therefore, the transaction price distribution will differ from the quote price distribution to the extent that the lowest quote differs from the average quote; this gap will be larger if the typical borrower obtains more quotes or if borrowers who obtain the same number of quotes, obtain the second one from a particularly low priced lender.

The first equation we estimate is

$$E[p_{l,m}^{quote}] = a^q + b^q BSI_m + c^q X_m + u_m \tag{6}$$

where the price expectation is taken using quote share weights, and X_m is a vector of county characteristics that might be associated with differential search rates. We have also estimated this equation using $E[p_{l,m}^{orig}]$, the price expectation taken using the origination share weights, as the dependent variable. Both regressions have also been re-estimated using the value of the BSI obtained from the log-linear equation 2.[29]

The results are shown in Table 12. The average quote and origination price in a market decline with BSI. This holds true for both the linear and log model, although the effects are stronger for the log model. Furthermore, for both models, the magnitude of decrease is somewhat larger for the origination price mean, though this difference is small.The reduction of the mean price quote associated with BSI has two components. The first is that financial institutions are less likely to operate in counties with consumers who are prone to collect more information; the second is that among the institutions operating in those counties, informed prospective borrowers are more likely to file applications with the lower cost ones. In the next section, we examine and discuss the first component more systematically. We briefly note here that when we use the average price of a lender in a county, $E[p_{l,m}^{lender}]$, as the dependent variable in equation 7, the decline of that price with BSI_m is small. Therefore, to the extent to which BSI leads to lower quotes, the driver is primarily differences in the probabilities with which borrowers file applications to lenders in different parts of the price distribution. Pure random sampling would not lead to lower quotes relative to the average price charged by lenders in a county: it would simply yield more-of-the-same type of quotes. This seems to suggest that borrowers rely on prior information from sources in their social network (or other sources) when they submit applications and this information results in a directed search focusing on lenders with lower mortgage rates.

The distribution of origination prices is even more responsive to BSI. One possibility for this responsiveness is that borrowers in markets with high search propensity file more applications, and as a result they obtain a better transaction rate holding the distribution of offers fixed. However, as discussed in the pre-

[29]The estimation is again done via GLS, implemented through iteratively re-weighted least squares. We have also estimated the model with instrumenting for HHI, as in the estimation of the equation (4).

ceding section, there is a nearly complete "crowd out" of the higher propensity of certain borrower-types to file more applications. The observed reduction in origination rates, then, must be an outcome of directed search: Informed individuals, when they file a second application, do so for an institution that has very competitive pricing.

In the second equation we estimate, we move beyond the central tendency in prices, and look at how the entire offer and transaction price distribution depend on underlying search propensities. In particular, we estimate the equation

$$Q[p_{l,m}^{quote}|\tau] = a^q + b^q BSI_m + c^q X_m + u_m \tag{7}$$

where $Q[p_{l,m}^{quote}|\tau]$ is the τ^{th} percentile of the quote distribution $F_m^A(x)$. Equation (7) is a quantile regression, but $Q[p_{l,m}^{quote}|\tau]$ is obtained from the corresponding quantile $F_m^A(x)$. Thus, this equation is not estimated via quantile methods, where observations from all quantiles are used as the dependent variable and the quantile check function is used to re-weight the objective appropriately to yield parameter estimates for the desired value of τ. Rather, the dependent variable is directly the τ quantile of the distribution, and the equation is estimated via linear regression, indeed as all the other regressions via GLS implemented using iteratively re-weighted least squares.

These regressions also have the same complement of explanatory variables as the mean regressions. However, in the results, reported in Table 13, the estimates of these other explanatory variables are omitted to conserve space. A higher BSI is associated with lower price deciles for both the quote and the origination price distributions, and the effects are stronger for the origination price distribution. Moreover, the effects are generally more significant for deciles around and below the median. By and large all locations have the same set of high priced lenders. Mid and low price lenders are more frequent and/or receive a disproportionate number of applications in high information environments.

5 Lenders' Market Presence

A financial institution's deposit rates and rates for other financial products will often differ across the localities that it operates in. However, mortgage rates of a given lender tend not to vary within states, and sometimes do not vary across states either. As noted earlier, the main reason is the fear of being accused of "redlining." Therefore, lenders cannot tailor mortgage rates to the local market conditions, including to the search intensity in a particular market. Their rates would reflect the competitiveness in the entire set of markets in which they operate, and would also reflect their operating costs and brand name (large lenders would be more prominent, would generate more traffic and could charge higher rates).

Whatever is the optimal rate of the lender, the institution can decide to be active or inactive in a particular market, i.e., it can decide whether to enter a market or stay out. A high search propensity area would be one where a lender can obtain smaller profits, all things equal. This would be particularly relevant for high price (high cost) lenders. They are the ones that increased search activity or increased information would tend to hurt disproportionately, as borrowers would be able to identify and choose lower cost alternatives. Therefore, high price/cost lenders should be particularly prone to stay out of high search areas, leading to lower prices but due to selection. However, mortgages are only part of the sales portfolio of lenders. Search induced competitiveness in the mortgage market would have strong effects in overall market competitiveness only if it is correlated with competitiveness in other products (other loans, deposit accounts, etc.). This is a strong, but plausible premise. Consumers who search intensively for mortgages and exchange information about their experience with each other are also prone to search intensively and exchange information for other financial products.

With the above discussion in mind, we investigate whether lenders' entry decision into a particular market is affected by the baseline search intensity of a certain market and the lender's rate. The lender's interest rate is (mostly) an exogenous object, determined primarily by decisions other than the participation decision in that particular market. Recall that BSI_m is also an exogenous object: it is the propensity to search and obtain information, and not the actual search in a market. We start by creating a lender-county level dataset as follows: 1) For each state s, we record all the lenders l active in state s, i.e. lenders that have at least one search record from state s in the HMDA data.[30] 2) For each active lender l, if it has at least one search record in county m (that belongs to state s), we assign the entry indicator $e_{l,m} = 1$; otherwise we assign $e_{l,m} = 0$. Let the total number of lenders that are active in state s be L_s, and let the total number of counties in that state be M_s. Then, in this dataset, there are $L_s \times M_s$ observations for state s. Approximately, 15% of the lender-county level observations in this constructed dataset have $e_{lm} = 1$. We then match the

[30] We also tried a looser criterion of entry using all applications instead of only "searches " (as defined in the data section) as an indicator of being active in a certain market. The results are similar, although the BSI effects on entry are weaker.

lenders in each state with their price, $p_{l,s}$, as obtained from the estimation of equation (5).[31]

We use this dataset to estimate, via a probit regression, the probability that a lender is active in market m as a function of that market's BSI, the lender's price, and a host of other market characteristics. Because we want the marginal effect of BSI to vary flexible with the lender's price, the specification we employ is a flexible spline with respect to $p_{l,s}$ and its interaction with BSI_m. In particular, we estimate:

$$\begin{aligned} Pr(e_{l,m}=1) &= \alpha+\beta BSI_m+\delta p_{l,s_m}+\gamma BSI_m\cdot p_{l,s_m}+\sum_{\lambda=1}^{\lambda=\Lambda}\delta_\lambda \max(0,p_{l,s_m}-p_\lambda) \\ &\quad +\sum_{\lambda=1}^{\lambda=\Lambda}\gamma_\lambda BSI_m\cdot\max(0,p_{l,s_m}-p_\lambda)+\phi X_m+\nu_{lm} \qquad (8) \end{aligned}$$

where $p_i(\lambda=1,2,...,\Lambda)$ are knots for the spline function, the subscript s_m refers to the state s where county m is located in, and X_m are the market characteristics we've used earlier in this paper.[32] The set of knots contains the 10th, 25th, 50th, 75th, and 90th percentiles of the rates in this lender-county level dataset.[33] Besides controlling for county-level market characteristics, we also include state fixed effects.[34]

The results of this analysis are reported in Table 14. More instructive is the impact of BSI_m on the probability that a lender is active in market m. In Figure 1, we plot the derivative of the entry probability with respect to BSI for different BSI percentiles as a function of a lender's price (lending rate) for the linear and log model, respectively. These figures contain a number of different piecewise linear curves for different values of BSI_m. Somewhat easier to read is Figure 2, which tries to capture the marginal effect of BSI_m on entry probabilities using a single line. We construct the lines in Figure 2 by first calculating the derivative of entry probability with respect to BSI for every lender-county observation and then run a spline regression with these derivatives as the dependent variable and the lender rates as independent variables, using the 9 deciles of rates as knots. For both sets of figures, we note that the derivative of the entry probability with respect to BSI_m is always negative, indicating that higher search intensity markets are associated with lower entry rates. The effect is stronger for high priced lenders, as indicated by the severe decline in the plotted lines between the 30th and 80th percentiles of bank rates, which correspond to lending rates of 4.1%

[31]Since we only have price data for loans that are securitized by Fannie, Freddie, or Ginnie, roughly 60% of the actual lender entries ($e_{l,m}=1$) in HMDA are matched to the corresponding lender prices. These lenders account for 78% of the originations in HMDA, and 97% of the mortgages in the Fannie, Freddie and Ginnie database.

[32]Note that for every lender p_{l,s_m} is actually the same for all markets in the same state.

[33]Another set of knots we tried contains all 9 deciles of the rates in this lender-county level dataset. The graphs are much spikier, but the general tenor of the results remains the same.

[34]The lender price is the only lender characteristic, because we want price to be the summary measure of all lender characteristics that affect its pricing structure. Ideally, we would like to have the expected price of the lender, but do not have enough lender characteristics to get reasonable estimates of it. Adding lender characteristics would remove some of the exogenous component of the observed price.

to 4.3%. Pricier lenders are more sensitive to information-induced market competitiveness than low priced lenders. At the highest price levels, the relationship between entry probabilities and BSI_m flattens at a low value. The most expensive of banks in a market probably obtain much of their revenue from unrelated sources; though their presence in a market is adversely affected by BSI_m, it is no more so than those banks that are just a bit less costly than them.

6 Robustness Checks

INCOMPLETE. We performed a number of variations of this analysis as robustness checks and the associated results are largely qualitatively similar to those reported, though differences are quantitatively material in some instances.

We repeat our analysis using only counties with less than 300 originations in 2014. These are "small" counties where entry and exit may be more important. The results are shown in Tables **??**-**??**.

Instead of using iteratively re-weighted least squares, we weigh our observations with the number of originations in the first stage regression and all price regressions, using analytic weights. We weigh our second stage regression with the standard errors of the estimated market fixed effects. The results are shown in Tables **??**-**??**.

We have added pre-approvals to the set of applications. Because pre-approval data is not available for every financial institution, we imputed pre-approvals for institutions that do not report them on a pro-rata basis. The results are essentially the same as those reported here.

We also tried running the first stage without any weights, incorporating pre-approvals, and varying the set of loan and property characteristics used. Our results remain mostly unchanged. We omit these tables for brevity.

7 Concluding Remarks

TO BE COMPLETED.

References

Alexandrov, Alexei and Sergei Koulayev, (2017). "No Shopping in the U.S. Mortgage Market: Direct and Strategic Effects of Providing Information," manuscript.

Allen, Jason, Robert Clark and Jean-François Houde, (2016). "Search Frictions and Market Power in Negotiated Price Markets", NBER working paper 19883.

Amel, D., Kennickell, A., Moore, B., (2008). "Banking market definition: evidence from the survey of consumer finances," Unpublished working paper. Finance & Economic Discussion Series, U.S. Federal Reserve Board.

Armstrong, Mark, John Vickers, and Jidong Zhou, (2009). "Prominence and Consumer Search," *RAND Journal of Economics*, vol. 40, pages 209-233.

Brown, Jacqueline Johnson, and Peter H. Reingen, (1987). "Social Ties and Word-of-Mouth Referral Behavior," *Journal of Consumer Research*, vol. 14, pages 350-362.

Brown, Zach, (2016). "An Empirical Model of Price Transparency and Markups in Health Care," manuscript.

Burdett, Kenneth and Kenneth L. Judd, (1983). "Equilibrium Price Dispersion," *Econometrica*, vol. 51, pages 955-969.

Calem, Paul S. and Leonard I. Nakamura, (1998). "Branch Banking and the Geography of Bank Pricing," *Review of Economics and Statistics*, vol. 80, pages 600-610.

Deltas, George and Zening Li, (2017). "Consumer Search and Price Discrimination in the Mortgage Loan Market", manuscript.

Gurun, Umit G., Gregor Matvos and Amit Seru, (2013). "Advertising Expensive Mortgages," Fama-Miller Working Paper; Kreisman Working Papers Series in Housing Law and Policy No. 10.

Hortacsu, Ali and Chad Syverson, (2003). "Product Differentiation, Search Costs, and Competition in the Mutual Fund Industry: A Case Study of S&P Index Funds," NBER working paper 9728.

Lacko, James and Janis Pappalardo, (2007). "Improving Consumer Mortgage Disclosures," Bureau of Economics Staff Report, U.S. Federal Trade Commission.

Lang, William W. and Leonard I. Nakamura, (1993). "A Model of Redlining," *Journal of Urban Economics*, vol. 33, pages 223-234.

Manning, Richard and Peter Morgan, (1982). "Search and Consumer Theory," *Review of Economic Studies*, vol. 49, pages 203-216.

Mondragon, John, (2015). "Household Credit and Employment in the Great Recession," manuscript.

Morgan, Peter and Richard Manning, (1985). "Optimal Search," *Econometrica* , vol. 53, pages 923-944.

Salop, Steven and Joseph Stiglitz, (1977). "Bargains and Ripoffs: A Model of Monopolistically Competitive Price Dispersion," *Review of Economic Studies*, vol. 44, pages 493-510.

Salz, Tobias, (2015). "Intermediation and Competition in Search Markets: An Empirical Case Study," manuscript.

Scharfstein, David and Adi Sundaram, (2014). "Market Power in Mortgage Lending and the Transmission of Monetary Policy," manuscript.

Sorensen, Alan T., (2000). "Equilibrium Price Dispersion in Retail Markets for Prescription Drugs," *Journal of Political Economy*, vol. 108, pages 833-850.

Stahl, Dale O. II, (1989). "Oligopolistic Pricing with Sequential Consumer Search," *American Economic Review*, vol. 79, pages 700-712.

Stigler, George T., (1961). "The Economics of Information," *Journal of Political Economy*, vol. 69, pages 213-225.

Varian, Hal R., 1980). "A Model of Sales," *American Economic Review*, vol. 70, pages 651-659.

Woodward, Susan E. and Robert E. Hall, (2012). "Diagnosing Consumer Confusion and Sub-optimal Shopping Effort: Theory and Mortgage-Market Evidence." *American Economic Review*, 102: 3249-3276.

Whyte, William H. (1954) "The Web of Word of Mouth," *Fortune* , vol. 50 (November), 140-143.

Wilde, Louis L. and Alan Schwartz, (1979). "Equilibrium Comparison Shopping," *Review of Economic Studies*, vol. 46, pages 543-553.

MAIN TABLES

Table 1: **Actions Taken for HMDA Applications**

Action Taken	Obs.	Percent
Loan originated	5,466,417	60.92
Application approved but not accepted	322,802	3.60
Application denied by financial institution	1,589,281	17.71
Application withdrawn by applicant	1,045,079	11.65
File closed for incompleteness	364,126	4.06
Pre-approval denied by financial institution	123,077	1.37
Pre-approval approved but not accepted (optional reporting)	62,966	0.70
Total	8,973,748	100.00

Table 2: **HMDA Data Selection Process**

	Selection Criteria	Observations remaining
(1)	Applications for first-lien, 1-4 family homes	8,973,748
(2)	Drop applications not categorized as searches	6,834,298
(3)	Drop if county code missing	6,787,902
(4)	Drop if applicant is not a natural person	6,700,772
(6)	After grouping	940,436

Table 3: **Borrower/Loan Characteristics Distribution from HMDA**

Characteristics	Obs.	Percent
Race		
White	5,189,597	77.45%
Asian	384,865	5.74%
Black or African American	382,671	5.71%
American Indian or Alaska Native	41,637	0.62%
Native Hawaiian or Other Pacific Islander	26,771	0.40%
Not Provided	675,231	10.08%
Gender		
Male	4,488,120	66.98%
Female	1,807,624	26.98%
Not Provided	405,028	6.04%
Ethnicity		
Not Hispanic or Latino	5,469,814	81.63%
Hispanic or Latino	591,839	8.83%
Not Provided	639,119	9.54%
Loan Purpose		
Home Purchase	3,621,387	54.04%
Refinancing	2,892,392	43.17%
Home Improvement	186,993	2.79%
Owner-occupancy Status		
Owner-occupied	5,951,496	88.82%
Not Owner-occupied	749,276	11.18%
Loan Type		
Conventional	4,953,252	73.92%
FHA-insured	975,801	14.56%
VA-guaranteed	611,602	9.13%
FSA/RHS	160,117	2.39%
Total	6,700,772	100%

Table 4: **HHI and Instruments**

Characteristics	Min.	Max.	Mean	Std. Dev.	Obs.
HHI (2014)	0.014	1	0.10	0.10	3,211
HHI (2007)	0	1	0.08	0.07	3,206
HHI (Difference)	0	0.15	0.02	0.02	3,206

HHI is 0 when there are no originations in that county for the entire year.

Table 5: **Discrete Loan Characteristics from GSEs and Ginnie Mae**

Characteristics	Obs.	Percent
Mortgage Securitizer		
Fannie Mae	1,675,173	43.37%
Freddie Mac	1,078,121	27.91%
Ginnie Mae	1,109,448	28.72%
Third Party Origination Flag		
Broker	373,867	9.68%
Correspondent	1,360,707	35.23%
Retail	2,128,168	55.09%
Loan Purpose		
Purchase	2,356,485	61.01%
Non-purchase	1,506,257	38.99%
Number of Units		
1 Unit	3,781,789	97.90%
2 Units	58,194	1.51%
3 Units	12,117	0.31%
4 Units	10,642	0.28%
Number of Borrowers		
1 Borrower	2,034,853	52.68%
2 Borrowers	1,811,900	46.91%
$\geq$ 3 Borrowers	15,989	0.41%
Total	3,862,742	100%
Owner-occupancy Status (GSE only)		
Owner-occupied	2,362,188	85.79%
Not owner-occupied	391,106	14.21%
Property Type (GSE Only)		
Single Family	1,786,628	64.89%
Non-Single Family	966,666	35.11%
Total (GSE Only)	2,753,294	100%

Table 6: **Continuous Loan Characteristics from GSEs and Ginnie Mae**

Characteristics	Min.	Max.	Mean	Std. Dev.	Obs.
Origination Rate (%)	1.75	7.2	4.29	0.46	3,862,742
Credit Score	300	850	729.13	56.02	3,862,742
Loan-to-value ratio	1	873	81.33	18.40	3,862,742
Debt-to-income ratio	1	65	35.49	9.94	3,862,742
Loan Amount (in 000s)	620.5	1873	200.64	112.21	3,862,742
Origination Term (in months)	84	480	324.76	71.38	3,862,742

Table 7: **County Characteristics from ACS**

Characteristics	Min.	Max.	Mean	Std. Dev.	Obs.
ln(Population)	6.05	16.12	10.28	1.45	3,211
ln(PercapitaIncome)	8.76	11.06	10.04	0.27	3,211
ln(OwneroccupiedUnits)	3.69	14.22	8.98	1.39	3,211
ln(RentUnits)	3.76	14.37	8.02	1.54	3,211
ln(MedianRent)	5.51	7.50	6.50	0.25	3,211
ln(MedianValue)	9.90	13.72	11.67	0.45	3,211
BachelorsPct	0.03	0.41	0.13	0.05	3,211
MinorityPct	0	0.96	0.17	0.17	3,211
WorkerPct	0.19	5.17	0.81	0.24	3,211
EmploymentPct	0.66	1	0.91	0.04	3,211
LaborForcePct	0.21	0.90	0.59	0.08	3,211

1. ACS provide data on 3,220 counties. Here we only sumarize the 3,211 counties that are matched with the HMDA data and have at least 1 origination in 2014.
2. The variable "WorkerPct" is the defined as the percentage of workers working in a specific county (possibly commuting from other counties) divided by the labor force number in that county.

Table 8: **Search and Group Characteristics**

	Linear Model		Log Model	
	Search/Origination		ln(Search/Origination)	
Caucasian	0.004***	(0.001)	0.020***	(0.001)
African American	0.017***	(0.002)	0.011***	(0.001)
Asian	-0.017***	(0.002)	-0.016***	(0.001)
Male	-0.057***	(0.002)	-0.048***	(0.001)
Female	-0.072***	(0.002)	-0.066***	(0.001)
Hispanic	0.003	(0.002)	-0.001	(0.001)
Non-hispanic	0.026***	(0.001)	0.033***	(0.001)
Purchase	0.010***	(0.001)	0.028***	(0.001)
Refinance	0.130***	(0.001)	0.111***	(0.001)
Owner-occupied	0.023***	(0.001)	0.029***	(0.001)
Conventional	0.003**	(0.001)	0.011***	(0.001)
FHA-insured	0.007***	(0.001)	0.006***	(0.001)
VA-guaranteed	-0.021***	(0.002)	-0.019***	(0.001)
Loan Bin 2	-0.020***	(0.001)	-0.016***	(0.001)
Loan Bin 3	-0.024***	(0.001)	-0.018***	(0.001)
Loan Bin 4	-0.009***	(0.001)	-0.007***	(0.001)
Income Bin 1	0.002	(0.002)	0.023***	(0.001)
Income Bin 2	-0.019***	(0.002)	0.005***	(0.001)
Income Bin 3	-0.025***	(0.002)	0.000	(0.001)
Income Bin 4	-0.021***	(0.002)	0.003***	(0.001)
Constant	1.201***	(0.003)	0.096***	(0.002)
Observations	812,446		812,446	
R^2	0.955		0.446	

Huber-White standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: **County Effect and BSI**

	Linear Model		Log Model	
		IV: M&A 2007		IV: M&A 2007
BSI	-0.869***	-0.875***	-0.975***	-0.972***
	(0.064)	(0.065)	(0.048)	(0.052)
HHI	-0.081***	-0.091***	-0.038***	-0.061***
	(0.010)	(0.031)	(0.005)	(0.017)
ln(Population)	-0.007	-0.013	-0.010	-0.012*
	(0.012)	(0.010)	(0.007)	(0.007)
ln(PercapitaIncome)	-0.027***	-0.030***	-0.017***	-0.019***
	(0.008)	(0.008)	(0.005)	(0.005)
ln(OwnerOccupiedUnits)	0.021**	0.028***	0.027***	0.030***
	(0.010)	(0.008)	(0.005)	(0.005)
ln(RentUnits)	0.004	0.004	0.005**	0.004*
	(0.004)	(0.004)	(0.002)	(0.002)
ln(MedianRent)	0.045***	0.047***	0.032***	0.032***
	(0.007)	(0.007)	(0.005)	(0.005)
ln(MedianValue)	0.022***	0.024***	0.021***	0.023***
	(0.003)	(0.003)	(0.002)	(0.002)
BachelorsPct	-0.019	-0.029	-0.026*	-0.031*
	(0.023)	(0.023)	(0.015)	(0.016)
MinorityPct	-0.027***	-0.025***	-0.031***	-0.029***
	(0.008)	(0.007)	(0.004)	(0.004)
WorkerPct	0.003	0.002	-0.001	-0.000
	(0.005)	(0.004)	(0.003)	(0.003)
EmploymentPct	-0.205***	-0.201***	-0.124***	-0.122***
	(0.032)	(0.031)	(0.021)	(0.021)
LaborForcePct	-0.067***	-0.064***	-0.038***	-0.035***
	(0.015)	(0.015)	(0.010)	(0.010)
Constant	0.970***	0.961***	-0.192***	-0.196***
	(0.059)	(0.060)	(0.038)	(0.040)
Observations	3211	3206	3211	3206
R^2	0.609	0.609	0.771	0.769

Huber-White standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 10: **Mortgage Rates and Loan Characteristics**

	Mortgage Rate	
Credit Score (/1000)	-1.458***	(0.003)
Original LTV (%)	0.230***	(0.001)
Debt-to-Income Ratio (%)	0.052***	(0.002)
Original Loan Amount (millions)	-0.674***	(0.002)
Original Term (months/1000)	4.591***	(0.002)
Purchase	-0.104***	(0.000)
Correspondent	0.036***	(0.001)
Retail	0.078***	(0.001)
Single Unit	-0.213***	(0.001)
Single Borrower	-0.000	(0.000)
Ginnie	-0.487***	(0.000)
Constant	4.189***	(0.003)
Observations	3,747,364	
R^2	0.665	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 11: **Constructed Price Distribution Summary Statistics**

Variable	Min.	Max.	Mean	Std. Dev.	Obs.
Quote Price Mean	3.921	4.491	4.230	0.064	3,203
Quote Price 10 pctl	3.656	4.491	4.063	0.092	3,203
Quote Price 20 pctl	3.689	4.491	4.113	0.090	3,203
Quote Price 30 pctl	3.797	4.491	4.152	0.090	3,203
Quote Price 40 pctl	3.797	4.491	4.190	0.087	3,203
Quote Price 50 pctl	3.815	4.631	4.228	0.082	3,203
Quote Price 60 pctl	3.921	4.631	4.263	0.075	3,203
Quote Price 70 pctl	3.921	5.127	4.298	0.071	3,203
Quote Price 80 pctl	4.039	5.127	4.336	0.066	3,203
Quote Price 90 pctl	4.135	6.120	4.385	0.083	3,203
Origination Price Mean	3.918	4.469	4.225	0.065	3,197
Origination Price 10 pctl	3.656	4.384	4.065	0.093	3,197
Origination Price 20 pctl	3.689	4.384	4.113	0.092	3,197
Origination Price 30 pctl	3.716	4.384	4.152	0.093	3,197
Origination Price 40 pctl	3.797	4.456	4.188	0.090	3,197
Origination Price 50 pctl	3.797	4.631	4.225	0.086	3,197
Origination Price 60 pctl	3.921	4.631	4.261	0.078	3,197
Origination Price 70 pctl	3.921	4.700	4.296	0.072	3,197
Origination Price 80 pctl	4.013	4.700	4.333	0.065	3,197
Origination Price 90 pctl	4.050	5.687	4.376	0.073	3,197

Table 12: **Quote & Origination Price Mean and BSI**

	Linear Model				Log Model			
	E[P^{quote}]	E[P^{quote}] (IV)	E[$P^{orig.}$]	E[$P^{orig.}$] (IV)	E[P^{quote}]	E[P^{quote}] (IV)	E[$P^{orig.}$]	E[$P^{orig.}$] (IV)
BSI	-0.197**	-0.148*	-0.226***	-0.207**	-0.462***	-0.438***	-0.560***	-0.538***
	(0.080)	(0.080)	(0.081)	(0.080)	(0.093)	(0.093)	(0.094)	(0.094)
HHI	-0.178***	-0.140***	-0.193***	-0.133***	-0.182***	-0.147***	-0.191***	-0.133***
	(0.022)	(0.037)	(0.022)	(0.036)	(0.022)	(0.037)	(0.022)	(0.038)
ln(Population)	0.095***	0.094***	0.100***	0.101***	0.092***	0.094***	0.103***	0.104***
	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
ln(PercapitaIncome)	0.079***	0.080***	0.076***	0.078***	0.089***	0.090***	0.089***	0.091***
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
ln(OwnerOccupiedUnits)	-0.098***	-0.098***	-0.103***	-0.102***	-0.098***	-0.099***	-0.105***	-0.105***
	(0.009)	(0.009)	(0.010)	(0.009)	(0.009)	(0.009)	(0.009)	(0.009)
ln(RentUnits)	-0.001	-0.000	-0.003	-0.004	0.000	-0.000	-0.003	-0.004
	(0.004)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.005)
ln(MedianRent)	0.121***	0.123***	0.124***	0.127***	0.116***	0.117***	0.118***	0.121***
	(0.008)	(0.009)	(0.009)	(0.009)	(0.008)	(0.009)	(0.009)	(0.009)
ln(MedianValue)	-0.032***	-0.031***	-0.031***	-0.030***	-0.032***	-0.032***	-0.033***	-0.032***
	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
BachelorsPct	-0.243***	-0.246***	-0.217***	-0.232***	-0.272***	-0.278***	-0.262***	-0.270***
	(0.028)	(0.029)	(0.028)	(0.030)	(0.029)	(0.030)	(0.029)	(0.030)
MinorityPct	0.014**	0.015**	0.009	0.009	0.012*	0.015**	0.008	0.009
	(0.007)	(0.007)	(0.007)	(0.007)	(0.006)	(0.007)	(0.007)	(0.006)
WorkerPct	-0.009*	-0.011**	-0.006	-0.009*	-0.009*	-0.011**	-0.007	-0.009*
	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
EmploymentPct	-0.037	-0.032	-0.030	-0.023	-0.040	-0.030	-0.030	-0.025
	(0.033)	(0.033)	(0.033)	(0.033)	(0.033)	(0.034)	(0.033)	(0.033)
LaborForcePct	-0.266***	-0.269***	-0.263***	-0.261***	-0.264***	-0.266***	-0.260***	-0.264***
	(0.018)	(0.018)	(0.019)	(0.019)	(0.018)	(0.018)	(0.019)	(0.019)
Constant	3.188***	3.149***	3.166***	3.107***	3.175***	3.140***	3.137***	3.073***
	(0.059)	(0.066)	(0.061)	(0.068)	(0.060)	(0.067)	(0.061)	(0.070)
Observations	3203	3198	3197	3192	3203	3198	3197	3192
R^2	0.339	0.337	0.339	0.343	0.346	0.341	0.354	0.348

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 13: **Quote & Origination Price Percentiles and BSI**

	10pctl	20pctl	30pctl	40pctl	50pctl	60pctl	70pctl	80pctl	90pctl
Quote Price (Linear Model)									
BSI	-0.154	-0.263	-0.576***	-0.620***	-0.454**	-0.405**	-0.597	-0.590	-0.256
	(0.184)	(0.215)	(0.184)	(0.204)	(0.218)	(0.205)	(0.363)	(0.471)	(0.243)
BSI (IV)	-0.131	-0.328*	-0.583***	-0.864***	-0.554*	-0.466*	-1.389	-1.390*	-0.585
	(0.173)	(0.196)	(0.181)	(0.326)	(0.283)	(0.252)	(0.914)	(0.778)	(0.589)
Origination Price (Linear Model)									
BSI	-0.154	-0.266	-0.449**	-0.612***	-0.464*	-0.348	-0.238	-0.040	-0.291
	(0.156)	(0.182)	(0.189)	(0.225)	(0.241)	(0.260)	(0.200)	(0.290)	(0.206)
BSI (IV)	-0.175	-0.272	-0.563**	-1.018**	-1.060**	-0.535	-0.467	-0.415*	-0.843*
	(0.169)	(0.186)	(0.221)	(0.404)	(0.436)	(0.388)	(0.289)	(0.235)	(0.493)
Quote Price (Log Model)									
BSI	-0.369*	-0.658**	-0.981***	-0.991***	-0.796***	-0.719***	-0.578*	-0.422	-0.241
	(0.218)	(0.273)	(0.244)	(0.298)	(0.271)	(0.250)	(0.329)	(0.423)	(0.262)
BSI (IV)	-0.366*	-0.727***	-1.050***	-1.241***	-1.032***	-0.933***	-1.270**	-1.083*	-0.561
	(0.211)	(0.261)	(0.248)	(0.411)	(0.272)	(0.264)	(0.605)	(0.602)	(0.563)
Origination Price (Log Model)									
BSI	-0.503***	-0.724***	-0.911***	-1.016***	-0.833***	-0.648**	-0.434*	-0.204	-0.259
	(0.190)	(0.228)	(0.241)	(0.335)	(0.299)	(0.316)	(0.242)	(0.322)	(0.243)
BSI (IV)	-0.464**	-0.741***	-0.935***	-1.507***	-1.155***	-1.067**	-0.887**	-0.911***	-0.745*
	(0.205)	(0.235)	(0.270)	(0.487)	(0.412)	(0.518)	(0.384)	(0.263)	(0.435)

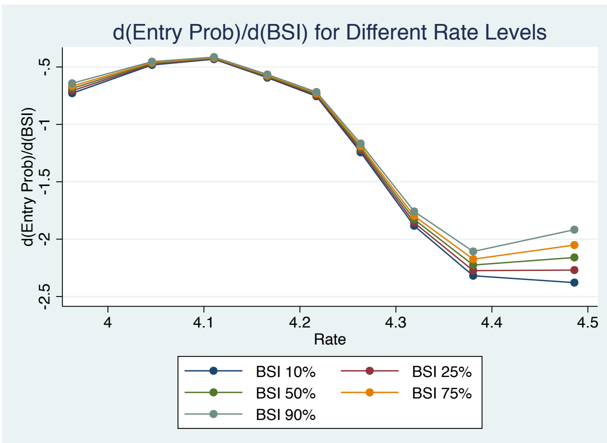
Huber-White standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 14: **Lender's Entry Decision (Stricter Criterion)**

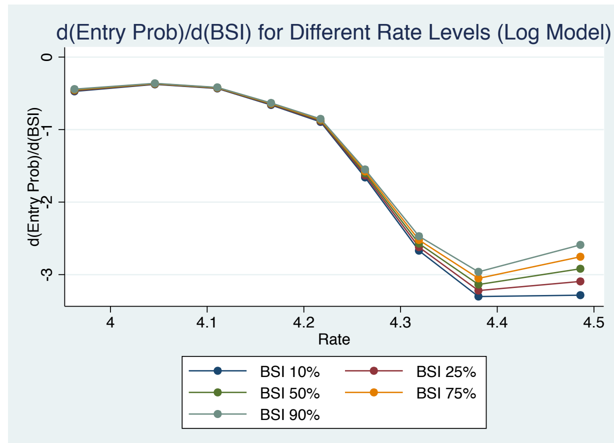
	Pr(entry=1)			
	Linear Model		Log Model	
BSI	30.924	(19.636)	20.481	(23.282)
rate	1.285***	(0.126)	1.591***	(0.529)
BSI · rate	-8.638*	(5.031)	-5.723	(5.966)
$max(0, r - r_{10})$	0.308	(0.252)	-0.353	(1.045)
$max(0, r - r_{25})$	0.448*	(0.254)	1.617	(1.043)
$max(0, r - r_{50})$	1.007***	(0.204)	3.328***	(0.830)
$max(0, r - r_{75})$	-6.276***	(0.202)	-8.332***	(0.822)
$max(0, r - r_{90})$	2.981***	(0.134)	0.603	(0.547)
BSI·$max(0, r - r_{10})$	25.666***	(9.797)	14.261	(11.728)
BSI·$max(0, r - r_{25})$	-24.167**	(9.578)	-19.556*	(11.646)
BSI·$max(0, r - r_{50})$	-19.323**	(7.528)	-31.400***	(9.265)
BSI·$max(0, r - r_{75})$	12.498*	(7.447)	26.371***	(9.203)
BSI·$max(0, r - r_{90})$	26.292***	(4.942)	33.985***	(6.128)
ln(Population)	-0.011	(0.025)	-0.018	(0.025)
ln(PercapitaIncome)	-0.004	(0.025)	0.012	(0.025)
ln(OwnerOccupiedUnits)	0.391***	(0.021)	0.397***	(0.021)
ln(RentUnits)	-0.007	(0.009)	-0.005	(0.009)
ln(MedianRent)	0.349***	(0.019)	0.326***	(0.020)
ln(MedianValue)	0.285***	(0.012)	0.289***	(0.012)
BachelorsPct	-0.212***	(0.067)	-0.181***	(0.066)
MinorityPct	-0.118***	(0.019)	-0.167***	(0.019)
WorkerPct	-0.116***	(0.010)	-0.117***	(0.010)
EmploymentPct	-0.916***	(0.086)	-0.871***	(0.086)
LaborForcePct	-0.036	(0.041)	-0.067	(0.041)
Constant	-12.346***	(0.528)	-13.559***	(2.072)
Observations	773,707		773,707	

Standard errors in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

FIGURES

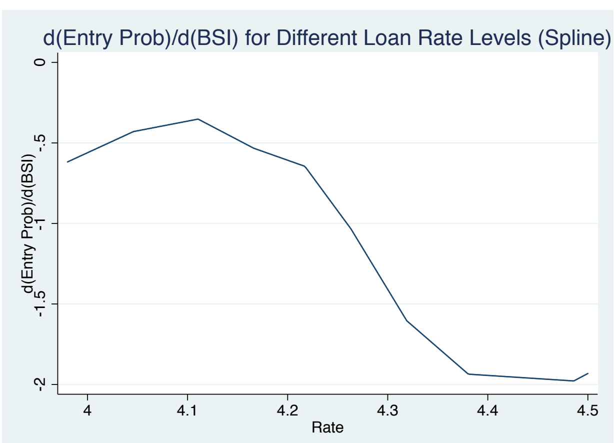


(a) Linear Model

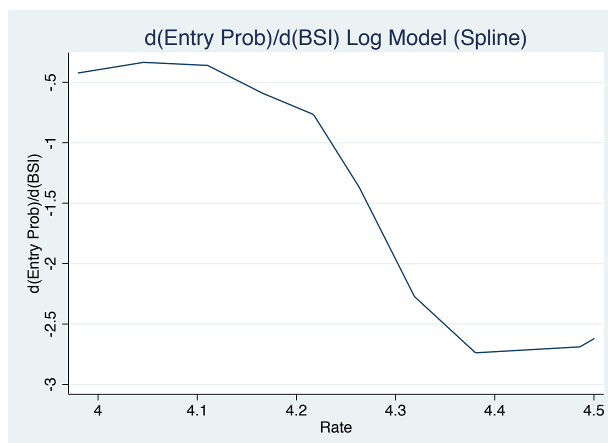


(b) Log Model

Figure 1: $\frac{\partial EntryProb.}{\partial BSI}$ for Different Rate Levels



(a) Linear Model



(b) Log Model

Figure 2: $\frac{\partial EntryProb.}{\partial BSI}$ for Different Rate Levels (Spline Regression)

APPENDIX TABLES

APPENDIX TABLES CURRENTLY REVISED.