Credit Constraints and Residential Sorting*

Edward Kung[†] Ralph Mastromonaco[‡]

September 28, 2015

Abstract

We study how credit constraints affect residential sorting. Using a novel dataset that includes information on both housing choice and the credit-worthiness of home buyers, we document that credit score influences housing choice for borrowers but does not affect the choices of non-borrowers, indicating that borrowers are credit-constrained. Using a discrete choice model of housing choice augmented with credit constraints, we show that ignoring credit constraints biases down willingness-to-pay estimates for various amenities by 2-6 percent. We estimate that the average willingness-to-pay for a 100 point increase in credit scores is \$17 per month, but there is substantial heterogeneity—as high as \$100 per month for some borrowers. Low-income borrowers are more likely than high-income borrowers to increase their consumption of housing amenities in response to a credit score increase. We estimate that credit constraints can explain 20 percent of the inequality in housing amenities.

JEL Codes: D10, R21

*We thank Matthew Kahn, Chris Timmins, Nic Kuminoff, and Maurizio Mazzocco for helpful comments.

[†]UCLA. ekung@econ.ucla.edu.

[‡]University of Oregon.

1 Introduction

Location is tied to a variety of amenities, such as access to high quality education, exposure to crime, and peers with which to form beneficial social interactions. Because of the high correlation between location and amenities, where a person chooses to live can have a large impact on economic outcomes.[1] Understanding residential sorting—the process by which households choose where to live—is therefore a key component to understanding the distribution of economic outcomes. Even if location does not affect long-run economic outcomes, understanding residential sorting is still important because of the intrinsic consumption value of local amenities.

One aspect of residential sorting that has largely been overlooked by the literature is the role of credit markets. Credit plays an important role in industrialized economies like the United States. Poor credit can restrict access to banking services, rental properties, and employment opportunities. Moreover, poor credit can increase borrowing costs for home and auto loans, or prevent access to loans altogether. This is of paramount concern for residential sorting, as greater access to credit can unlock the opportunity to purchase a home in a neighborhood with less crime or better schools, while poor credit does the opposite. Heterogeneity in access to credit would therefore imply heterogeneity in the access to housing opportunities. Since access to credit tends to be correlated with income and other demographic characteristics, credit constraints can have important distributional consequences for residential sorting.

In this paper, we study the role of credit constraints in residential sorting. We have three main contributions. First, we provide evidence that credit constraints indeed have a causal effect on residential sorting. Second, we demonstrate that ignoring credit constraints can bias one's estimates for consumer preferences over housing amenities, and therefore bias a model's implications about residential sorting. Finally, we provide the first (to our knowledge) esti-

[1] The extent to which location affects outcomes is under debate. Experimental studies have found little evidence that neighborhoods affect outcomes for adults (Katz et al. (2001); Oreopoulos (2003)), but Chetty and Hendren (2015) find large neighborhood exposure effects for children.

mates of the housing market benefits to having a better credit score, and the first estimates of the impact of credit constraints on residential inequality. We find that there are significant benefits to increasing one's credit score in the housing market, and we find that credit constraints can explain a substantial fraction of residential inequality.

The data that we use for this paper comes from the housing market of the Atlanta-Sandy Springs-Marietta, GA metropolitan statistical area in 2011. The unique feature of the data is that we observe not only the housing choices of individuals in the data, but also the creditworthiness of the individual. Such a dataset combining information from both housing and credit markets has previously been unavailable to researchers, and the lack of data may explain the relative dearth of studies related to credit markets and residential sorting. We view the construction and analysis of the data as another key contribution of the paper.

We demonstrate in two ways that credit constraints have a causal effect on residential choice. First, we show that the price of home purchased by a buyer is increasing and discontinuous in the buyer's credit score. Lenders use credit scores to evaluate the riskiness of borrowers, and a higher credit score is generally associated with lower borrowing costs and greater access to credit. *Ceteris paribus*, a borrower with a higher credit score can therefore afford to purchase a more expensive home, which is reflected in the increasing relationship between purchase price and credit score. One may be concerned that credit scores are proxying for unobserved preferences, but it is unlikely that preferences would be discontinuous in credit scores, whereas it is well-established that lenders treat credit scores in a discontinuous fashion (Keys et al. (2010)). Second, we show that the increasing relationship between purchase price and credit score holds only for borrowers but not for non-borrowers. If credit scores were merely a proxy for unboserved preferences, then the relationship between credit scores and purchase price would continue to hold even for buyers who did not need to borrow. We show that these results are robust to the inclusion of a number of demographic controls, and that they hold for housing amenities as well as just purchase price.

To investigate the bias arising from ignoring credit constraints, and to estimate the impact of credit constraints on residential sorting, we develop a residential sorting model with credit constraints. The model contributes to the existing literature by taking a standard model of residential sorting (i.e. Bayer et al. (2007)) and introducing heterogeneous borrowing costs, as well as the possibility that some homes and neighborhoods are simply out of reach for some buyers. In the model, households choose to buy one house out of a set of possibilities. Household decisions are influenced by their preferences over housing amenities and the price of each home. Credit constraints affect the mortgage interest rate, introducing heterogeneity in the financing costs across households and homes, and may even limit the set of possible homes that a household has access to (through limits on the loan amount).

We estimate the model with and without credit constraints. We find that ignoring credit constraints biases downwards the estimated willingness-to-pay for various amenities by 2 to 6 percent. Intuitively, ignoring credit constraints biases preference estimates downwards because it makes individuals appear more price-sensitive than they actually are. In the data, there are some individuals who do not choose expensive homes with good amenities, not because they are unwilling to pay for the amenities, but rather because they cannot afford it at all. We also find that ignoring credit constraints biases the estimated relationships between preferences and demographic characteristics. Intuitively, access to credit is correlated with demographic characteristics, like income. If credit constraints are ignored, then the impact of residential sorting based on credit scores is mistakenly attributed to sorting based on demographic characteristics.

In the last section of the paper, we use our estimates from the model augmented with credit constraints to study a number of counterfactual questions. First, we ask, what are the economic benefits to increasing one's credit score in the housing market? For each individual in our data, we simulate what their residential outcomes would have been if their credit score was unilaterally increased by 100 points. We find that the average willingness-to-pay (i.e. compensating variation) for such an increase is, on average, $17 per month, which is not a small amount. However, there is substantial heterogeneity across households. For some households, the benefit of a 100 point credit score increase is over $130 per month. Because we account for residential sorting, these estimate differ from the "engineering estimate" of the benefit of an increased credit score which only considers reduced borrowing costs. Importantly, we find that increasing an individual's credit score also results in the purchase of more housing amenities. The price of the purchased homes goes up by an average of 8 percent, school quality goes up by 0.04 standard deviations, square-footage goes up by 3 percent, and the income of the chosen neighborhood goes up by 2 percent. Again, there is substantial heterogeneity in the increase in amenities resulting from an increase in credit scores. Importantly, we find that the correlation between income and amenities goes down when credit scores are increased. The results imply that credit constraints account for some of the income-based residential sorting across neighborhoods.

To investigate more formally the contribution of credit constraints to income-based sorting and other measures of residential inequality, we use the model to simulate residential outcomes if credit constraints were removed. In this counterfactual, borrowing costs are homogenized across individuals and it is assumed that any individual may borrow any amount, thus eliminating the effect of credit constraints on choice sets. We find that when credit constraints are removed, the variance in the choice of residential amenities is reduced, as is the correlation between amenities and income. The results imply that credit constraints can explain about 20 percent of the standard deviation in residential amenities consumption, and about 6 percent of the correlation with income. We should emphasize at the outset that these are partial equilibrium results—the simulations are carried out under the assumption that house prices and the housing supply do not change. Nevertheless, the suggest show that the impact of credit constraints on residential inequality is potentially very large.

Related Literature

This paper is related to two broad strands of literature. First, our work is related to the literature studying the general interactions between credit and housing markets. Second, the paper is related to the vast body of work estimating the value of local amenities, and more broadly, of demand estimation and the valuation of non-market and public goods.

There is a growing body of research that empirically demonstrates the relationship between credit markets and housing markets. Most of the literature has focused on the relationship between credit markets and house prices (Adelino et al. (2012); Favara and Imbs (2015); Kung (2015); Anenberg et al. (2015)). Some papers consider the effect of liquidity constraints on residential mobility (Chan (2001); Ferreira et al. (2010)), though these papers are more focused on the effect of negative equity in locking a homeowner into his or her current home. Theoretically, it is understood that liquidity constraints can affect the quality of house that individuals purchase (Stein et al. (1995); Ortalo-Magne and Rady (2006); Bajari et al. (2013)), with implications for how house prices respond to various shocks, but few papers have directly examined the role of credit constraints in where households choose to live—particularly in regard to their consumption of local amenities. The only other paper we are aware of is a recent study by Ouazad and Rancière (2013). Using MSA-level bank liquidity as an instrumental variable for credit availability, they show that the reduction in credit constraints in the first half of the 2000s had an impact on residential mobility. Interestingly, they argue that the reduction in credit constraints actually increased racial segregation, because poorer white households were likely to move out of racially integrated neighborhoods as a response.

Our result that credit constraints can bias preference estimates has important implications for the literature on demand estimation and the evaluation of non-market goods. The fundamental models used in non-market valuation have not changed substantially in past years. Most of the work is built on the hedonic model of Rosen (1974) or on discrete choice models such as in Berry et al. (1995) and Bayer et al. (2007). These revealed preference methods have been used to evaluate a variety of non-market goods, including local public school quality (Black (1999); Bayer et al. (2007), neighborhood crime (Linden and Rockoff (2008)), environmental quality (Gamper-Rabindran and Timmins (2012); Walsh (2007)), and neighborhood demographics (Bajari and Kahn (2005)). Much of the research has focused on implementing novel methods for controlling for unobserved amenities at the neighborhood or house level. In contrast, less attention has been paid to unobserved heterogeneity at the individual level, which is what we focus on, through the impact of credit constraints.

Lastly, there has been a push to improve the performance of non-market valuation methods. One strand has focused on adding dynamics. Bayer et al. (2011) and Bishop and Murphy (2011) include dynamic, forward looking households to value neighborhood attributes. Mastromonaco (2015) and Caetano (2012) use dynamic models to estimate the willingness to pay for school quality. A second strand, more related to the current study, focuses on modeling of the choice set. Banzhaf and Smith (2007) examine the sensitivity of willingness to pay estimates from discrete choice models to different assumptions on budget constraints and choice affordability. Manzini and Mariotti (2013) implement a model of bounded rationality and allow decision makers to vary by their level of attention. The goal of these papers is to demonstrate an inadequacy of the current models to accurately reflect the nature of the data generating process, specifically attacking the assumption of myopic, unconstrained decision makers. Similarly, our study seeks to push the literature forward by relaxing assumptions, specifically that of perfect credit markets and unlimited borrowing.

2 Data

The core of our data comes from Dataquick, a real-estate consulting company. In the Dataquick data, each record is a real estate transaction, collected from property deeds and tax assessor data from across the United States. It contains the universe of all property transactions for the counties and years in which data was collected.

We begin with Dataquick data on all single-family-home and condo sales in 2011 for the Atlanta-Sandy Springs-Marietta, GA, metropolitan statistical area (MSA). For each property transaction, Dataquick records the name of the buyer, the street address of the property, the purchase price, and the size of any loans taken out against the property at the time of purchase. The data also contains a variety of property characteristics, such as the square-footage and the number of bedrooms and bathrooms.

We supplement the Dataquick data with information about local demographics and amenities. We focus on three neighborhood attributes for each house: the income-level of the Census-block-group (CBG), the share of population that is white in the CBG, and the quality of the local public elementary school. To obtain data on CBG income and race, we use data from the 2010 Census. To obtain data on local public school quality, we map each house to its local elementary school using attendance zone maps from the School Attendance Boundary Information System, or SABINS. We then obtain quality data for each of these schools from GreatSchools, an information provider for public school quality. Our measure of school quality focuses on the elementary level score on the Criterion-Referenced Competency Tests (CRCT). The quality variable is then constructed by taking the fraction of 4th grade students that meet or exceed the state standard for both reading and math, and average them together. The quality variable therefore varies from 0 to 100, with 100 indicating that all 4th grade students in the school met the state standards on both reading and math.[2]

To obtain information about the credit characteristics of our buyers, we supplement this data with credit reports provided by Experian, one of the three major credit reporting companies. The buyer in each Dataquick record is matched to his or her December 2010 Experian credit report. For each individual in their database, Experian maintains a list of known addresses, which, combined with address information from Dataquick, facilitates matching. Af-

[2] We have experimented with using other measures, such as the 5th grade proficiency rate, or the raw scores of reading and math. The qualitative results are not much changed.

ter matching, Experian returns to us a de-identified dataset merged with two key variables: the buyer's "Vantage" score and "Income Insight" score. Vantage Score is Experian's proprietary credit scoring model, and is comparable to the better known FICO score. Income Insight is an individual-level income estimate provided by Experian. Experian did not disclose to us the precise methodology that they used to construct this estimate, but they reveal that it is estimated off verified income information, such as W2's, as well as credit information.

In addition to these, Experian also provides us with the buyer's age, gender, occupation, education, and marital status. Like Income Insight, these demographics are individual-level estimates based on Experian's proprietary methods. Importantly, there is variation in these demographics within our lowest levels of geography (CBG), so these are individual-based measures and not simply geographic aggregates. As an example of how Experian arrives at these measures, for education they report: "Information is compiled from self-reported surveys, derived based on occupational information, or calculated through the application of a predictive model."

Table 1 shows summary statistics for the Experian variables using our matched dataset. It also shows the number of matched and unmatched observations. The match rate is quite good at 71 percent. Table 2 shows summary statistics for housing and neighborhood attributes in the matched sample. The average buyer in our data has a credit score of about 720, though there is a significant fraction with lower credit scores as well (28 percent of the sample has a credit score below 680). The average annual income among our matched buyers is $97,000, the average age of the buyer is 44, 53 percent have 4-year college degrees or higher, and 78 percent are married. The average price that the buyers paid for their homes was $230,000, and the average home was 2,300 square feet with 3 bedrooms and 3 bathrooms. The average school quality is 87, the average CBG income is $93,000, and the average CBG percent white is 67 percent.

How representative is our matched sample compared to the original unmatched sample? Unfortunately, the answer is "not very." Table 3 compares housing and neighborhood attributes in the matched and unmatched samples. The houses in the matched sample are of clearly better quality than the unmatched sample. They are higher priced, are in better school zones and neighborhoods, and are larger. This suggests that our unmatched buyers are predominantly unmatched because they lack credit information. If they were unmatched predominantly due to random inaccuracies such as spelling errors, then we would expect the matched sample to look similar to the unmatched sample. As a result, the results in this paper should be understood to apply only to the population of home buyers who have credit reports.

3 Evidence that credit constraints affect housing choice

Figure 1 shows the spatial distribution of housing transactions in our matched sample, along with elementary-school attendance boundaries. Figure 2 shows the spatial distribution of school quality, house prices, buyer income, and buyer credit scores. The geographic distribution of these variables show that higher-income and higher-credit-score individuals sort into neighborhoods with higher house prices and better local public schools. These patterns show that both income and credit scores matter for housing and neighborhood choice. Income and credit scores can affect housing choice through a variety of channels. One possibility is that the buyer's income and credit score affect the amount and interest rate at which he is able to borrow, and therefore affects which homes he is able to afford. This is the credit channel that we are interested in. But besides the credit channel, it may also be the case that higher-income and higher-credit-score buyers simply prefer larger homes for reasons completely unrelated to the credit market.

In this section, we will isolate the credit channel from unobserved preference heterogeneity. We will argue that credit score is *not* correlated with unobserved housing preference, and is therefore a good measure of the access to credit that an individual has. We make this argument using two approaches. First, we show that housing choice is discontinuous in credit scores. Because unobserved preferences are unlikely to be discontinuous in credit scores, and because we know that credit scores are treated discontinuously by lenders in the mortgage market (c.f. Keys et al. (2010)), any discontinuity in the relationship between housing choice and credit scores should reflect constraints in the mortgage market rather than unobserved preferences. Second, we show that housing choice is *not* correlated with credit scores among cash buyers. If credit scores are correlated with unobserved preferences, then we would expect correlations between credit score and housing choice, even among non-borrowers.

Figure 3 plots average house prices by buyer credit score, using our matched sample. Figure 3 shows that house price is generally increasing in credit score, which is consistent with both the credit channel and other interpretations. Figure 3 also shows that house price is discontinuous in credit scores, with the two most apparent discontinuities occuring at 720 and 770. Figures 4 and 5 confirm that the discontinuities are driven by lending; they show, respectively, that the discontinuities are reflected in loan amounts, and that there is a discontinuity in the ease of obtaining a jumbo loan at 770 credit score. We do not know exactly how banks are operating in our dataset, but we do know from the existing literature that lenders use credit score thresholds in determining credit availability. The discontinuity in house prices is therefore consistent with the credit channel. It is less consistent with unobserved preferences because it is unlikely that preferences would be discontinuous in borrower credit scores.[3]

To further verify that the discontinuities in house price with respect to credit scores are not driven by underlying preferences, we run the following

[3] One possible way in which preferences may be discontinuous in credit scores is if home buyers are able to manipulate their credit scores, and some buyers are more likely than others to manipulate their score to just above the threshold. This seems unlikely, however, as credit scoring algorithms are quite complex and closely guarded by the companies that produce them. It does not seem plausible that a large enough number of buyers are able to manipulate their credit scores to a great enough precision so as to generate a preference-based discontinuity around a credit score threshold.

regressions:

$$\begin{aligned}\log p_i &= \alpha_1 1\left[Score_i > 720\right] + \alpha_2 1\left[Score_i > 770\right] \\ &\quad + \alpha_3 Score_i + \alpha_4 Score_i^2 + \alpha_5 Score_i^3 + \mathbf{X}_i \beta + \epsilon_i \end{aligned} \tag{1}$$

Here, the log transaction price of buyer i is regressed against indicators for whether the buyer's credit score is above 720 and 770, a third-order polynomial of credit score, and a vector of demographic variables which includes income, marital status, education, and whether the buyer works a white-collar job. The results are presented in column 1 of Table 4. The results show that even when we control for demographic variables that are likely to affect preferences, the discontinuities with respect to credit score continue to be present.

The discontinuities show that the credit channel is present and affects housing choice. But we are also interested in the extent to which the (non-discontinuous) relationship between credit scores and housing choice is driven by the credit channel as opposed to unobserved preferences. To address this question, we note that if credit scores are simply proxying for unobserved preferences, then the correlation between credit score and housing choice should be present even among cash buyers. If, on the other hand, the correlation between credit score and housing choice is being driven by differential access to credit, then the correlation should not exist for the buyers who did not need to borrow. In columns 2-4 of Table 4, we repeat regression (1) for different subsamples of the data. In column 2, we perform the regression for buyers who borrowed at least 50 percent of their house price when purchasing. These are the buyers who needed to borrow, and we see from Table 4 that credit score continues to matter for the price of the home they buy. Both the discontinuities and the continuous relationship are present. In column 3, we perform the regression for buyers who borrowed at less than 50 percent of their house price when purchasing. These are buyers who needed to borrow, but borrowing was relatively less important for them. Table 4 shows that the discontinuities are no longer present for these borrowers, and the relationship between credit scores and purchase price is fairly weak. Finally, in column 4 we perform the regressions for cash buyers. Here, we see that credit scores do not appear to matter at all. In contrast to credit scores, the demographic variables continue to be statistically and economically significant in all three subsamples, and the magnitudes of the coefficients remain very stable throughout each specification. Overall, the results support the interpretation that credit scores reflect access to credit and influence housing choice through the credit channel, while the demographic variables reflect preferences that continue to influence housing choice even for non-borrowers.[4]

To further reinforce this interpretation, we perform F-tests of joint significance for the credit-score variables and for the demographic variables, in each of the subsamples in Table 4. For credit scores, we test the null hypothesis that the coefficients on the indicator variables are jointly zero, and that the third-order polynomial is different when evaluated at 500 than at 830. For the demographic variables, we test the null hypothesis that the coefficients are jointly zero. Table 5 reports the results from these tests of joint significance. Table 5 shows that in the full sample, and in the sample of high-LTV borrowers, both credit scores and demographic variables are jointly significant. By contrast, in the low-LTV and cash-buyer samples, only the demographic variables are jointly significant. The results suggest that buyers with higher credit scores purchase more expensive homes because they have better access to credit in the mortgage market. Either banks are willing to lend them more, or at a lower interest rate, or both.

In addition to transaction price, we are also interested in whether credit scores affect the actual housing and neighborhood amenities purchased. We repeat regression (1) using school quality, square footage, number of bathrooms, CBG percent white, and CBG income on the left-hand-side. The estimated coefficients are reported in Table 6 and the associated F-tests are reported in Table 7. For most of the attributes, the credit-score variables are significant in the full sample but not in the cash-buyer sample. Demographic variables

[4]We do not suggest that demographic variables, such as income or occupation, cannot also affect access to credit. Rather, we argue that once we control for demographic variables, credit scores only reflect access to credit, and do not capture unobserved preferences which influence housing choice.

are jointly significant for all attributes and in all samples. As with the price regressions, these results suggest that differential access to credit, as reflected through credit scores, affect the consumption of various housing and neighborhood amenities. We note that overall, the results are weaker when considering actual amenities as opposed to transaction price. This is to be expected since the purchase price of a home is a good single-index summary of the overall quality of a house, whereas individuals face many different tradeoffs when considering the consumption of specific amenities. Because of these tradeoffs, the theoretical relationship between credit access and specific amenities is more complicated than the relationship between credit access and house price. For this reason, Tables 4 and 5 are our preferred specifications for establishing the presence of the credit channel and the independence of credit-scores to unboserved preferences.

4 A model of housing choice with credit constraints

Having established a credit channel in influencing housing choice, we now estimate a model of housing choice that incorporates credit constraints. The purpose of the model is twofold. First, by estimating the model with and without credit constraints, we demonstrate the biases that result from ignoring credit constraints in preference estimation. Second, the model will inform us on the degree to which credit constraints affect residential sorting in reality. We are especially interested in the impact of credit scores on where individuals choose to live, and the access to housing and neighborhood amenities that this entails.

The model is an extension of the discrete-choice model of residential sorting in Bayer et al. (2007) (henceforth BFM). In the model, households $i = 1, \dots, N$ choose over houses $h = 1, \dots, H$, where $N = H$ to reflect the feature of the data that each buyer purchases one house. The price of each house is p_h, and it is assumed that all buyers are able to purchase the house at that price (though some buyers may have to borrow more to purchase it, and some buyers may be excluded from purchasing it if they cannot obtain a large enough loan).

The utility that household i obtains from choosing house h is given by:

$$V_{ih} = \beta_i \mathbf{x}_h - \alpha_i cost_{ih} + \theta_h^b + \xi_h + \epsilon_{ih} \tag{2}$$

$\mathbf{x}_h$ is a vector of characteristics for house h, which here includes the square-footage of the house, the number of bathrooms, the local public school quality, the CBG income, and the CBG percent white. $cost_{ih}$ is the annualized cost of purchasing the house, which, in a departure from BFM, we allow to vary across households i to reflect heterogeneity in borrowing costs. θ_h^b is the level of unobserved neighborhood amenities that we capture by assigning each house to a school attendance zone boundary and including boundary fixed effects.[5] ξ_h is an unobserved quality level of house h, which we assume is uncorrelated with $\mathbf{x}_h$ and θ_h^b. Finally, ϵ_{ih} is an idiosyncratic taste-shifter that affects household i's preference for house h. We assume that ϵ_{ih} is i.i.d. across households and houses, and that it is distributed according to a type-1 extreme value distribution.

$\beta_{i,k}$ is household i's marginal utility for housing attribute k, and α_i is household i's marginal utility over other consumption. By indexing the coefficients β_i and α_i by i, we allow for preference heterogeneity across households. Following BFM, we model preference heterogeneity as follows:

$$\beta_{i,k} = \gamma_k^0 + \sum_{\ell=1}^{L} \gamma_k^\ell z_i^\ell \tag{3}$$

$$\alpha_{i,k} = \rho^0 + \sum_{\ell=1}^{L} \rho^\ell z_i^\ell \tag{4}$$

[5]The usage of boundary fixed effects is discussed further in Black (1999) and Bayer et al. (2007). The central idea is to isolate the effect of school quality from other neighborhood amenities by comparing the prices of homes on two sides of a school attendance zone boundary. Each house in the data is thus assigned to the nearest attendance zone boundary, and boundary fixed effects are included in the model to control for the level of neighborhood amenities around the boundary. Kuminoff et al. (2010) and Kuminoff and Pope (2014) have argued that boundary fixed effects models are an effective, theoretically congruent econometric strategy for controlling for unobserved heterogeneity at the neighborhood level.

where z_i^ℓ is a demographic characteristic of household i. The L demographic characteristics we include are income, indicators for middle and old age (45-65 and above 65, respectively), an indicator for marital status, and an indicator for whether i is college educated. We do not allow household i's credit score to affect its preferences, conditional on the demographic variables. We believe this is a reasonable exclusion given the results in Section 3.

We depart from traditional models of discrete choice in two major ways. First, we allow for heterogeneity in the cost to purchasing a home, which is driven by heterogeneity in the interest rates that households face. The annualized cost for household i to purchase house h is given by:

$$cost_{ih} = r_{ih} \left(p_h - d_i\right) + r^f d_i \tag{5}$$

where r_{ih} is the interest rate that household i would have to pay when borrowing to pay for house h. We assume that r_{ih} depends on factors such as household i's credit score and the size of the loan, and we will explain in detail how r_{ih} is constructed in Section 4.1. d_i is the downpayment that household i has available to put towards the purchase of a house.[6] r^f is the return on savings available to the household.[7] The annualized cost of purchasing the home is therefore the interest paid on the borrowing necessary to purchase the home, plus the opportunity cost of the downpayment.

The second way in which we depart from traditional models is that we introduce heterogeneity in the choice set available to each household. Suppose without any loss of generality that the houses are ordered such that $p_1 < \ldots < p_H$ (i.e. in order of increasing prices). Furthermore, for each household i and home h, let ϕ_{ih} be the probability that h is the most expensive home that

[6] In BFM, the annualized cost is given by rp_h, where r was set at 5 percent. The BFM model therefore does not capture the effect of interest rate heterogeneity, nor the effect of heterogeneity in downpayments.

[7] We assume that $r^f < r_{ih}$ for all i and h to reflect the fact that households are generally not able to save at the same rate they must borrow at. When this is the case, households will put as much downpayment towards the house as they have available, and so we assume that the downpayment we observe in the data is the maximum downpayment that the household could have made towards any purchase.

household i could afford.[8] Limits on the choice set arise from limits to the amount of borrowing that households have access to. For example, it is very unlikely that a household making $50,000 a year with limited wealth and a low credit score will be able to borrow enough to purchase a million dollar home. The difficulty for borrowers of certain characteristics to obtain loans of a certain size are reflected in the choice set probabilities, ϕ_{ih}. We explain in detail how we model ϕ_{ih} in Section 4.2.

For notational convenience, let us now write:

$$V_{ih} = \lambda_{ih} + \delta_h + \epsilon_{ih} \tag{6}$$

where

$$\begin{aligned} \lambda_{ih} &= \sum_{k=1}^{K} \left(\sum_{\ell=1}^{L} \gamma_k^{\ell} z_i^{\ell} \right) x_{h,k} - \alpha_{i,k} cost_{ih} & (7) \\ \delta_h &= \sum_{k=1}^{K} \gamma_k^0 x_{h,k} + \theta_h^b + \xi_h & (8) \end{aligned}$$

λ_{ih} is the portion of V_{ih} that varies across both households and houses, while δ_h is the portion of V_{ih} that varies only across houses. δ_h can be thought of as the mean utility for house h in the population. Given this notation, the probability (over ϵ_{ih} and over potential choice sets) that household i chooses house h is:

$$P_{ih} = \sum_{q=h}^{H} \phi_{iq} \frac{\exp(\lambda_{ih} + \delta_h)}{\sum_{j=1}^{q} \exp(\lambda_{ij} + \delta_j)} \tag{9}$$

The parameters of the model can be estimated by maximum likelihood. The log likelihood of the data is simply given by:

$$LL = \sum_{i=1}^{N} \sum_{h=1}^{H} 1\left[h_i^* = h\right] \log P_{ih} \tag{10}$$

where h_i^* is the choice that household i is observed to have made in the data.

[8] In BFM and most other models of discrete choice, all houses (or alternatives) are assumed to be inside every household's choice set. In our notation, this would imply that $\phi_{iH} = 1$ for all i.

Direct estimation of the parameters is computationally difficult because there are many boundary fixed effects to estimate—one for each attendance zone boundary, 1,011 in total.

Fortunately, the structure of the choice probabilities (9) suggest a two-step approach that is also undertaken in Berry et al. (1995) and Bayer et al. (2007). In the first step, the parameters in λ_{ih} are estimated along with the mean-utilities δ_h for each house. Given a guess of the parameters governing λ_{ih}, the mean-utiltiies δ_h are computed using a contraction mapping. Equation (9) implies that δ_h satisfies a fixed-point equation, the sample analogue of which is:

$$\delta_h' = \delta_h + \log s_h - \log\left(\frac{1}{N}\sum_{i=1}^{N}\phi_{iq}\frac{\exp\left(\lambda_{ih}+\delta_h\right)}{\sum_{j=1}^{q}\exp\left(\lambda_{ij}+\delta_j\right)}\right) \tag{11}$$

Here, $\log s_h$ is the log-market-share of house h, which in our application is $1/N$. For the classical case in which all choices are available to all households ($\phi_{iH} = 1$ for all i), Berry et al. (1995) prove that equation (11) is indeed a contraction mapping, and therefore there is a unique solution for δ_h given λ_{ih}. Their proof may be extended to show that even for general values of ϕ_{ih}, equation (11) remains a contraction mapping and admits a unique solution.[9] The first step of the estimation is therefore to estimate the parameters in λ_{ih} by maximum likelihood, at each iteration computing the mean-utilities δ_h by contraction-mapping.

In the second step, the estimated values of δ_h from the first step are regressed against $\mathbf{x}_h$ and boundary-fixed-effects. From this regression, the the mean marginal utilities over housing attributes, γ_k^0, are estimated. OLS estimation of the second step is consistent because ξ_h is assumed to be independent of $\mathbf{x}_h$ and θ_h^b. After the second step, we will have estimated all the structural parameters of the model, and will be able to perform counterfactual exercises such as evaluating the willingness-to-pay for an increased credit score, and the effect that relaxing credit constraints has on housing choice.

[9]The proof is given in Appendix A.

4.1 Interest rate heterogeneity

In order to estimate the model, we need to know the interest rates that households face, not only for the loans they actually took out, but also for the loans they could have taken. Because this is a counterfactual, we will need to construct a model for interest rates as a function of borrower and loan characteristics. Unfortunately, our main dataset does not contain information about interest rates, so we turn to another data source to estimate the model. The data we use is the Fannie Mae Single Family Loan Level Dataset. This is a database of loan originations made publicly available by Fannie Mae. The dataset is representative of all the 30-year fixed rate, single family loans securitized by Fannie Mae. Because Freddie Mac and Fannie Mae were by far the dominant investors in the mortgage market in 2011, and both operate on fairly similar business practices, we believe that the Fannie Mae data is fairly representative of the mortgage market in 2011.[10]

Using the Fannie Mae data, we found that the following model predicted mortgage interest rates best:[11]

$$\begin{aligned} r_{ih} &= \tau_0 + \tau_1 \frac{\exp\left(\tau_2\left(\tau_3 - score_i\right)\right)}{1+\exp\left(\tau_2\left(\tau_3 - score_i\right)\right)} \\ &\quad + \tau_4 1\left[score_i > 680\right] + \tau_5 1\left[score_i > 720\right] \\ &\quad + \tau_6 1\left[LTV_{ih} > 0.8\right]\left(LTV_{ih} - 0.8\right) \end{aligned} \tag{12}$$

Here, the interest rate is non-linear in credit scores, with the slope being steepest at medium levels of credit scores. Interest rate is discontinuous in credit scores at 680 and 720.[12] The interest rate is also allowed to be increasing in loan-to-value ratio when the loan-to-value ratio crosses 80 percent.[13]

[10]We discuss two caveats at the bottom of this section.

[11]We have tried alternative specifications, such as including credit-scores linearly, using different variables, etc. We found that the model in (12) produced the best estimates.

[12]Interestingly, we did not find a discontinuity at 770, as we might expect from the results in Section 3. We return to this point later.

[13]This is an accurate reflection of the data, and is driven by rules governing Fannie Mae's lending. Fannie Mae guarantees mortgage debt up to 80% LTV, and for loans with higher

Equation (12) is estimated by non-linear least squares. The estimation results are reported in Table 8. Figure 6 plots the actual average interest rate by 10-point credit score bins against the fitted values. Figure 6 shows that the non-linear specification for interest rates in (12) is a good reflection of actual interest rates in the data. Interest rates are generally decreasing in credit score, and they are most sensitive to marginal changes to credit score in the 650 to 750 range.

A concern with our approach is that there are two segments of the mortgage market for which Fannie Mae data may not be representative. The first segment is the market for low credit score borrowers. Fannie Mae follows underwriting standards which restrict their lending to risky borrowers, and this is reflected in the distribution of credit scores in the Fannie Mae data relative to our data. In the Fannie Mae data, for example, the 25th percentile of credit scores is 729 while in our data it is 693. In defense of our approach, it is likely that low credit-score borrowers face higher interest rates from private investors than they do from Fannie Mae. This is because Fannie Mae, as one of the government sponsored enterprises, carries an implicit government guarantee on its credit obligations. This would make the interest rate profile with respect to credit-scores even steeper than what we estimate, which would only serve to reinforce our qualitative results.

The second segment of the mortgage market for which Fannie Mae data is not representative is the market for jumbo loans. Fannie Mae faces regulatory restrictions which prohibit it from securitizing loans of greater size than $417,000. There are therefore very few jumbo loans present in the Fannie Mae data. In our data, jumbo loans comprise a small, though not insignificant, fraction of the data (13 percent). Because of the lack of jumbo loans in the Fannie Mae data, we are unable to estimate the interest rate premium on jumbo loans. Traditionally, the jumbo loan premium has been estimated at somewhere between 20 to 50 basis points. For our results presented in Section

LTV it requires private mortgage insurance (PMI) or else a higher interest rate. The interest rate premium with respect to LTV is computed using loans without PMI. (We see which loans have PMI, but not what the actual mortgage insurance rate was.)

5, we assume a 20 basis point jumbo premium, applied additively to equation (12).[14]

4.2 Heterogeneity in choice sets

Besides interest rate heterogeneity, we also allow credit constraints to affect household decisions through their effect on choice sets. A difficulty in modeling choice sets is that they are not observed by the researcher. We only see the actual decision that each household made; we do not see all the choices they could have potentially made. To model the choice sets, we therefore take a conservative, non-parametric approach based on production frontier estimation. In frontier estimation, the problem at hand is to estimate the maximum amount of output achievable by a given amount of inputs. It is an appropriate analogy for thinking about credit availability, because we are concerned with estimating the maximum house price that a household could afford (output), given its characteristics (inputs).[15]

Defining H_i as the most expensive house that household i could afford ($\phi_{ih} = 1$ if $h = H_i$). We assume:

$$H_i = \max_j h_j^* \text{ s.t. } inc_j \leq inc_i,\ d_j \leq d_i,\ score_j \leq score_i \tag{13}$$

In words, our approach assumes that the most expensive house that i can afford is the most expensive house that someone observably worse than i (on three dimensions that lenders care about—income, downpayment, and credit score) actually did purchase. The idea behind this approach is that if a person j could borrow an amount L_j, then any person who is observably better than j on all dimensions that lenders care about can also borrow the amount L_j. In frontier estimation, this is the "free disposal" assumption. If, for every type of observably identical borrower, there is some probability that the borrower wishes to maximize its borrowing, then our method will correctly estimate the

[14] We have also estimated the model assuming a 50 basis point jumbo premium and found that our results do not change qualitatively. [We still need to do this for this draft].

[15] The frontier estimation approach has been applied fruitfully to mortgage data in Anenberg et al. (2015).

maximum borrowing amount for each borrower type.

Figure 7 plots the average house price frontier by 10-point credit score bins, computed using our main dataset from Atlanta. Unsurprisingly, the house price frontier is increasing in credit score. There appears to be a discontinuity in credit availability at score 770, which together with the interest rate discontinuities at 680 and 720, is consistent with the results from Section 3. Interestingly, although Figure 6 shows that interest rates are most sensitive to credit scores for medium score ranges, Figure 7 implies that actual credit availability is more sensitive at high and low credit score ranges.

Our approach assumes that lenders only use income, downpayment, and credit score to set maximum borrowing amounts. We believe this is a close approximation to reality, but there may still be other factors that lenders consider that we do not observe—such as whether the borrower has a pre-existing relationship with the lender. An advantage of our approach is that it is very conservative in the face of possible ommitted variables. In our method, if we see a low credit score borrower obtain a large loan amount, we assume that every borrower with higher credit score could also obtain this loan amount. But in reality, the low score borrower who obtained the large loan amount may be unobservably better. Our approach will therefore tend to overestimate the amount of credit availability to all borrower types. We will therefore understate the effect of credit constraints on housing choice, and so our results in Section 5 may be thought of as lower bounds. If we chose a more aggressive approach to modeling credit availability, we would find even stronger effects.

4.3 Discussion on the sources of identification

Intuitively, the preference parameteres in the model are identified through the observed correlations between purchased amenities and demographic characteristics. Conditional on house prices, if one observes positive sorting between income and school quality, then this suggests that higher income individuals have a higher preference for school quality than lower income individuals. Conditional on amenities, sorting by income across the price of homes identifies the relationship between income and the marginal utility of consumption. Residual sorting unexplained by price and observed amenities identifies the unobserved vertical quality of the homes. Neighborhood-level unobserved amenities are controlled for through the use of boundary fixed effects. Identification in discrete choice models are discussed more thoroughly in Berry et al. (1995); Bajari and Benkard (2005); Bayer et al. (2007).

5 Results

5.1 Naive estimates

Before estimating the full model with credit constraints, it is useful to first consider what the estimation results would be if credit constraints are ignored. These are the results that we would obtain from a traditional approach that does not consider heterogeneity in borrowing costs or credit availability. In this "naive" model, we assume homogeneity in interest rates, assigning the same interest rate to every individual based on the average interest rate computed from (12). The average interest rate was thus 5.07%. In the naive model, we also assume that every house is in every individual's choice set.

The estimation results for the naive model are reported in Table 9. Columns 1-5 report the estimates of γ_k^ℓ for housing attributes k and demographic characteristics ℓ. These are the coefficients determining the marginal utility of housing attributes, and how they vary by individual characteristics. Column 6 reports the estimates of ρ^ℓ, the marginal utility of consumption and how it varies by characteristics. The results are very reasonable. Starting from column 6, the results show that the marginal utility of consumption is positive, but decreasing in income. The marginal utilities over housing attributes are also positive, indicating that school quality, square footage, bathrooms, percent white, and neighborhood income are all goods. In most cases, the marginal utility over these attributes is increasing in income, indicating that higher income individuals generally have a higher willingness-to-pay for housing amenities than lower-income individuals. The marginal utltiies imply that the average willingness-to-pay for a 10-point increase in school quality is about \$50 per month. For the other attributes, we estimate a willingness to pay of: \$87/month for an additional 10 percent square feet; \$124/month for an additional bathroom; \$27/month for a 10 percentage point increase in neighborhood share white; \$23/month for a 10 percent increase in neighborhood income. These estimates are also reported in Panel A of Table 12.

5.2 Credit constrained estimates

We now present the estimation results from the full model with credit constraints. Table 10 reports the coefficient estimates and Table 11 reports the differences between the naive and the credit constrained models. As in the naive model, we continue to estimate a positive marginal utility for attributes which is generally increasing in income, and a positive marginal utility of consumption which is decreasing in income. However, the credit constrained model explains the data much better. Based on Akaike information criterion, the relative likelihood of the "naive" model to the "credit constrained" model is close to zero.

Column 6 of Table 11 shows that one of the key differences between the naive estimates and the credit constrained estimates is in the marginal utility of consumption. In the credit constrained model, we estimate a lower marginal utility of consumption overall, and a smaller relationship between the marginal utility of consumption and income. Intuitively, ignoring credit constraints causes us to overestimate the marginal utility of consumption because it makes individuals appear more price-sensitive than they actually are. In the data, there are some individuals who do not choose expensive houses with good amenities, not because they are unwilling to pay for the amenities, but rather because they cannot afford it at all (i.e. no bank is willing to lend them the requisite funds for such a purchase). Ignoring credit constraints also causes us to overestimate the relationship between income and the marginal utility of consumption because income is positively correlated with credit scores, and therefore negatively correlated with borrowing costs. If we ignore the effect of credit scores on borrowing costs, we underestimate the costs that low score borrowers pay to own their homes, and overestimate the costs of high score borrowers.

Table 12 summarizes our willingness-to-pay estimates across the two models, and for different subgroups of our sample. Panel A of Table 12 shows that, for the overall population, ignoring credit constraints will bias down our willingness-to-pay estimate for all housing attributes. The magnitude is modest, ranging from miniscule (for bathrooms) to 3 percent (for neighborhood racial composition and square-footage). Panels B and C show the willingness-to-pay estimates for college-educated, married individuals of age less than 45, with \$55,000 and \$116,000 annual income respectively (these are the 25th and 75th income percentiles in our data). Panels B and C show that willingness-to-pay for amenities is underestimated by the naive model across all attributes and borrower types, but there is heterogeneity in which estimates are most affected. For the low-income individual, ignoring credit constraints causes the most bias (as much as 6-7 percent) in the estimates for square-footage and neighborhood income. For the high-income individual, ignoring credit constraints causes the most bias in the estimates for school quality and neighborhood racial composition. Which attributes are more or less biased for what kinds of individuals is not easy to predict *a priori*. The magnitude of the bias depend on the characteristics of housing supply in the city. For example, in Atlanta, there are a number of low-priced houses in good school zones that even highly constrained individuals may have access to. However, there are not as many low-priced homes that are large and also in high-income neighborhoods. For this reason, ignoring credit constraints for highly constrained households biases our willingness-to-pay estimates for square-footage and neighborhood income moreso than for school quality.

The magnitude of the bias from ignoring credit constraints appears modest. This is a comforting result which suggests that existing estimates of the willingness-to-pay for various amenities, such as school quality, are not far off the mark. A caveat to this statement is that we have opted for a very conservative approach to modeling credit availability. It is likely that our method overestimates credit availability for many households in our data. A more aggressive (and perhaps realistic) approach would suggest a stronger bias from ignoring credit constraints. Our reported biases may therefore be interpreted as lower bounds for the effects of ignoring credit constraints on preference estimates.

Although the effect of ignoring credit constraints on willingness-to-pay estimates are modest, the effects on residential sorting are not. For example, the 2.2 percent bias in willingness-to-pay for a 10 point increase in school quality represents 13 percent of the standard-deviation in willingness-to-pay for school quality in the population. Since sorting across neighborhoods is more affected by relative preferences than by absolute preferences, the results imply that ignoring credit constraints can significantly alter the model's implications for sorting behavior. Figure 8 confirms this reasoning. The figure shows the expected choice of school quality by income for married, college-educated individuals under the age of 45, with 744 credit score and $15,000 downpayment (the population medians). The figure shows that the implied sorting behavior is much stronger in the naive model than in the credit constrained model. Ignoring credit constraints causes us to underpredict the choice of school quality among low-income individuals, and overpredict the choice of school quality among high-income individuals (holding credit characteristics constant). This pattern is similar for the other housing attributes as well.

5.3 The value of a credit score

The results suggest that credit scores have a significant impact on residential sorting through the cost of borrowing and through the availability of credit. Using our model, we can assess the benefits to increasing one's credit score. We consider the outcomes that each individual in our data would obtain if he or she *alone* were to receive an exogenous 100-point increase to credit score. Because in each counterfactual, only one individual receives a credit score shock, we assume that it has no effect on equilibrium house prices.

We first compute the willingness-to-pay for a 100 point increase in credit score. The willingness-to-pay is the amount by which annualized consumption must be reduced such that the individual is indifferent between receiving the credit score increase and having no change in circumstances. It is therefore a measure of compensating variation. As is appropriate, the counterfactual is simulated under the estimated parameters from the credit constrained model. Figure 9 plots each individual's willingness-to-pay for a 100 point credit score increase, reported in dollars-per-month, against their original credit score. The figure shows that the benefits are highest for medium credit-score borrowers, consistent with the fact that interest rates are most sensitive to credit scores in the medium score range. The benefits to a credit score increase can also be very large for some low-score borrowers. These are the borrowers whose choice sets are expanded significantly by the increase in credit scores. The average willingness-to-pay for a 100 point increase in credit scores is \$17 per month, and the maximum is \$131 a month. This exercise shows that the housing market benefits to increasing one's credit score can be substantial.

Figure 10 considers the effect of increasing one's credit score on residential sorting by income. Panels A through F of Figure 10 plot the expected increase in housing attributes (including price) that would result from a 100 point increase in credit score, for each individual in the data. The expected increase in housing attributes are plotted against income. Table 13 summarizes the results from Figure 10. The results show that the largest effect of increasing credit scores is in allowing households to purchase more expensive homes. On average, increasing credit score by 100 points increases the price of the chosen home by 8%. The average effect of increasing credit scores on actual amenities consumed is slightly more modest, suggesting that in addition to consuming more of the observed amenities, individuals who experience a credit score increase also consume more unobserved amenities, which are captured by the boundary fixed effects and the unobserved quality term ξ_h. For all the observed attributes, including price, the change induced by the credit score increase is negatively correlated with income, suggesting that increasing the credit scores of low income households has a greater effect on their consumption of housing amenities than for high-income households.

5.4 How much of urban residential inequality can be explained by credit constraints?

These results are interesting because they show that low-income households are disproportionately affected by constraints in the mortgage market. An important question to ask, therefore, would be how much of urban residential inequality can be explained by differential access to credit? Using the model, we can simulated the distribution of residential outcomes under the baseline, credit constrained model and compare it to the distribution of outcomes under a counterfactual model with no credit constraints. In the counterfactual model, we assume that the borrowing cost is 5.07% for all individuals, regardless of their credit score or loan amount, and we assume that all houses are in all individuals' choice sets—that is, every household can borrow enough to be able to purchase every house.

Table 14 reports the results from this exercise. Panel A reports the standard deviation in expected amenities chosen across individuals, and Panel B reports the correlation of the amenities chosen with income. On average, across all five measured amenities, credit constraints can explain about 20 percent of the standard deviation in the baseline model, and about 8 percent of the correlation with income. The results suggest that differential access to credit is a significant driver of urban residential inequality.

We must emphasize that these are partial equilibrium results, because we simulate the counterfactual model under the assumption that there are no changes to house prices. In reality, if credit constraints were removed, the demand for some houses would increase and the demand for others would decrease. Prices would have to adjust in response. *A priori*, it is not clear how large the price responses must be in the new equilibrium. Another general equilibrium effect that we ignore is re-sorting based on neighborhood demographics. If credit constraints were completely removed, this could change the demographic composition of some neighborhoods, leading some to become relatively more or less attractive. It is likely that this re-sorting behavior would only amplify the reduction in inequality first triggered by the removal of credit constraints, because as neighborhoods become more homogeneous, there is less of an incentive to sort on neighborhood demographics.

6 Conclusion

This paper demonstrates the role that credit constraints play in shaping urban residential outcomes. Our key findings are that credit constraints are a significant driver of residential sorting, that households (especially those of lower socioeconomic status) can benefit substantially from increased access to credit, and that credit constraints can explain a significant fraction of the inequality in residential outcomes.

Our results shed light on many observable phenomena in housing and credit markets. For example, anyone who watches television or listens to radio can attest to the many commercials advertising credit reporting services. In the popular press, one can also find many articles about how to improve one's credit score. These phenomena all suggest that credit scores are an important factor in peoples' decision-making, and to our knowledge we are the first to estimate the economic benefits from improving one's score. In the mortgage market, banks often advertise their downpayment, income, and credit score requirements for receiving a mortgage. When rates are low and credit standards are loose, borrowers are encouraged to purchase a home or refinance, consistent with our finding that credit constraints are an important factor in the decision of what to buy. Our results on credit constraints and urban inequality are also consistent with U.S. government policy that uses the mortgage market to increase housing affordability and reduce discrimination (the Home Mortgage Disclosure Act is an example of such a policy).

We refrain from drawing any policy conclusions from the results in this paper. First, we have not considered the impact to bank balance sheets of increasing credit accses to risky borrowers, or of eliminating the use of credit scores in setting interest rates. This is an important offsetting factor to the benefits of increasing access to credit. Second, we have only considered counterfactual experiments in which house prices remain fixed. This is appropriate when considering a small-scale policy that affects only a small subset of the population, but is inappropriate for evaluating the effects of large scale changes. Studying the general equilibrium effects of increasing credit access on residential outcomes is an important area for further study.

References

Adelino, Manuel, Antoinette Schoar, and Felipe Severino, "Credit supply and house prices: evidence from mortgage market segmentation," Technical Report, National Bureau of Economic Research 2012.

Anenberg, Elliot, Aurel Hizmo, Edward Kung, and Raven Molloy, "The Effect of Mortgage Credit Availability on House Prices and Investment: Evidence from a Frontier Estimation Approach," 2015. Working Paper.

Bajari, Patrick and C. Lanier Benkard, "Demand Estimation with Heterogeneous Consumers and Unobserved Product Characteristics: A Hedonic Approach," *Journal of Political Economy*, 2005, *113* (6).

_ and Matthew Kahn, "Estimating Housing Demand with an Application to Explaining Racial Segregation in Cities," *Journal of Business and Economic Statistics*, 2005, *23* (1), 20–33.

_ , Phoebe Chan, Dirk Krueger, and Daniel Miller, "A Dynamic Model of Housing Demand: Estimation and Policy Implications," *International Economic Review*, 2013, *54* (2), 409–442.

Banzhaf, H Spencer and V Kerry Smith, "Meta-analysis in model implementation: choice sets and the valuation of air quality improvements," *Journal of Applied Econometrics*, 2007, *22* (6), 1013–1031.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan, "A Unified Framework for Measuring Preferences for Schools and Neighborhoods," *Journal of Political Economy*, 2007, *115* (4), 588–638.

— **, Robert McMillan, Alvin Murphy, and Christopher Timmins**, "A Dynamic Model of Demand for Houses and Neighborhoods," July 2011. NBER Working Paper No. 17250.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–890.

Berry, Steven T., "Estimating Discrete-Choice Models of Product Differentiation," *RAND Journal of Economics*, 1994, *25* (2), 242–262.

Bishop, Kelly C and Alvin D Murphy, "Estimating the willingness to pay to avoid violent crime: A dynamic approach," *The American Economic Review: Papers and Proceedings*, 2011, *101* (3), 625–629.

Black, Sandra E., "Do Better Schools Matter? Parental Valuation of Elementary Education," *Quarterly Journal of Economics*, 1999, *114* (2), 577–599.

Caetano, Gregorio, "Neighborhood Sorting and the Valuation of Public School Quality," Working Paper, University of Rochester September 2012.

Chan, Sewin, "Spatial lock-in: Do falling house prices constrain residential mobility?," *Journal of urban Economics*, 2001, *49* (3), 567–586.

Chetty, Raj and Nathaniel Hendren, "The Impacts of Neighborhoods on Intergenerational Mobility: Childhood Exposure Effects and County-Level Estimates," 2015. Working Paper.

Favara, Giovanni and Jean Imbs, "Credit Supply and the Price of Housing," *American Economic Review*, 2015, *105* (3), 958–92.

Ferreira, Fernando, Joseph Gyourko, and Joseph Tracy, "Housing busts and household mobility," *Journal of urban Economics*, 2010, *68* (1), 34–45.

Gamper-Rabindran, Shanti and Christopher Timmins, "Does Cleanup of Hazardous Waste Sites Raise Housing Values? Evidence of Spatially Localized Benefits.," *Journal of Environmental Economics and Management*, 2012, *Forthcoming.*

Katz, Larry F., Jeffrey R. Kling, and Jeffrey B. Liebman, "Moving to Opportunity in Boston: Early Results of a Randomized Mobility Experiment," *Quarterly Journal of Economics*, 2001, *116* (2), 607–654.

Keys, Benjamin J., Tanmoy Mukherjee, Amit Seru, and Vikrant Vig, "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," *Quarterly Journal of Economics*, 2010, *125* (1), 307–362.

Kuminoff, Nicolai V and Jaren C Pope, "Do Capitalization Effects for Public Goods Reveal the Public's Willingness to Pay?," *International Economic Review*, 2014, *55* (4), 1227–1249.

_ , Christopher F Parmeter, and Jaren C Pope, "Which hedonic models can we trust to recover the marginal willingness to pay for environmental amenities?," *Journal of Environmental Economics and Management*, 2010, *60* (3), 145–160.

Kung, Edward, "The Effect of Credit Availability on House Prices: Evidence from the Economic Stimulus Act of 2008," Research Report, UCLA 2015.

Linden, Leigh and Jonah E. Rockoff, "Estimates of the Impact of Crime Risk on Property Values form Megan's Laws," *The American Economic Review*, 2008, *98* (3), 1103–1127.

Manzini, Paola and Marco Mariotti, "Stochastic Choice and Consideration Sets," Working Paper 2013.

Mastromonaco, Ralph, "A Dynamic General Equilibrium Analysis of School Quality Improvements," Technical Report, University of Oregon Working Paper 2015.

Oreopoulos, Philip, "The Long-Run Consequences of Living in a Poor Neighborhood," *Quarterly Journal of Economics*, 2003, *118* (4), 1533–1175.

Ortalo-Magne, Francois and Sven Rady, "Housing market dynamics: On the contribution of income shocks and credit constraints," *The Review of Economic Studies*, 2006, *73* (2), 459–485.

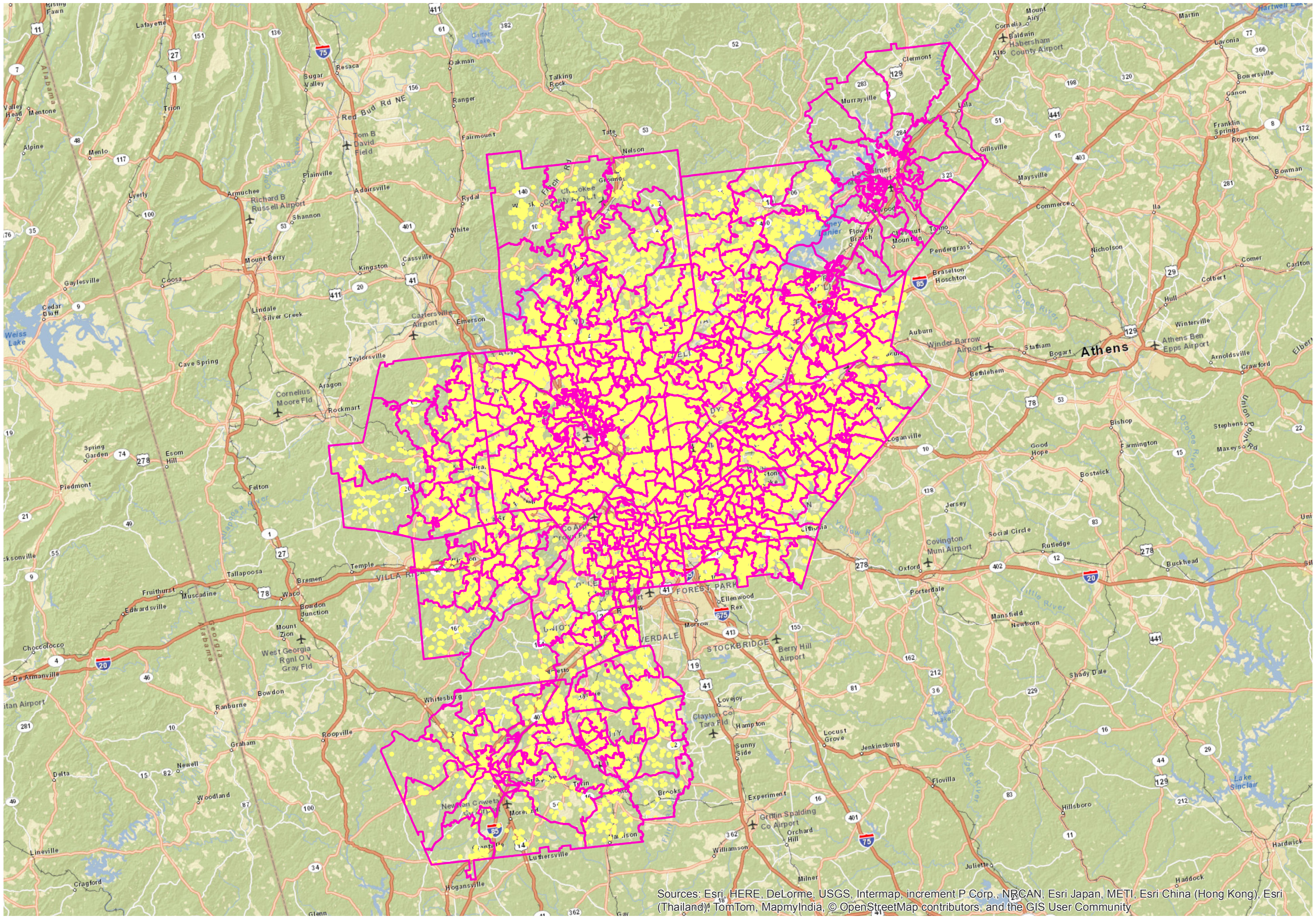
Ouazad, Amine and Romain Rancière, "Credit Standards and Segregation," 2013. Working Paper.

Rosen, Sherwin, "Hedonic Prices and Implicit Markets: Product Differentiation in Perfect Competition," *Journal of Political Economy*, 1974, *82* (1).

Stein, Jeremy C et al., "Prices and Trading Volume in the Housing Market: A Model with Down-Payment Effects," *The Quarterly Journal of Economics*, 1995, *110* (2), 379–406.

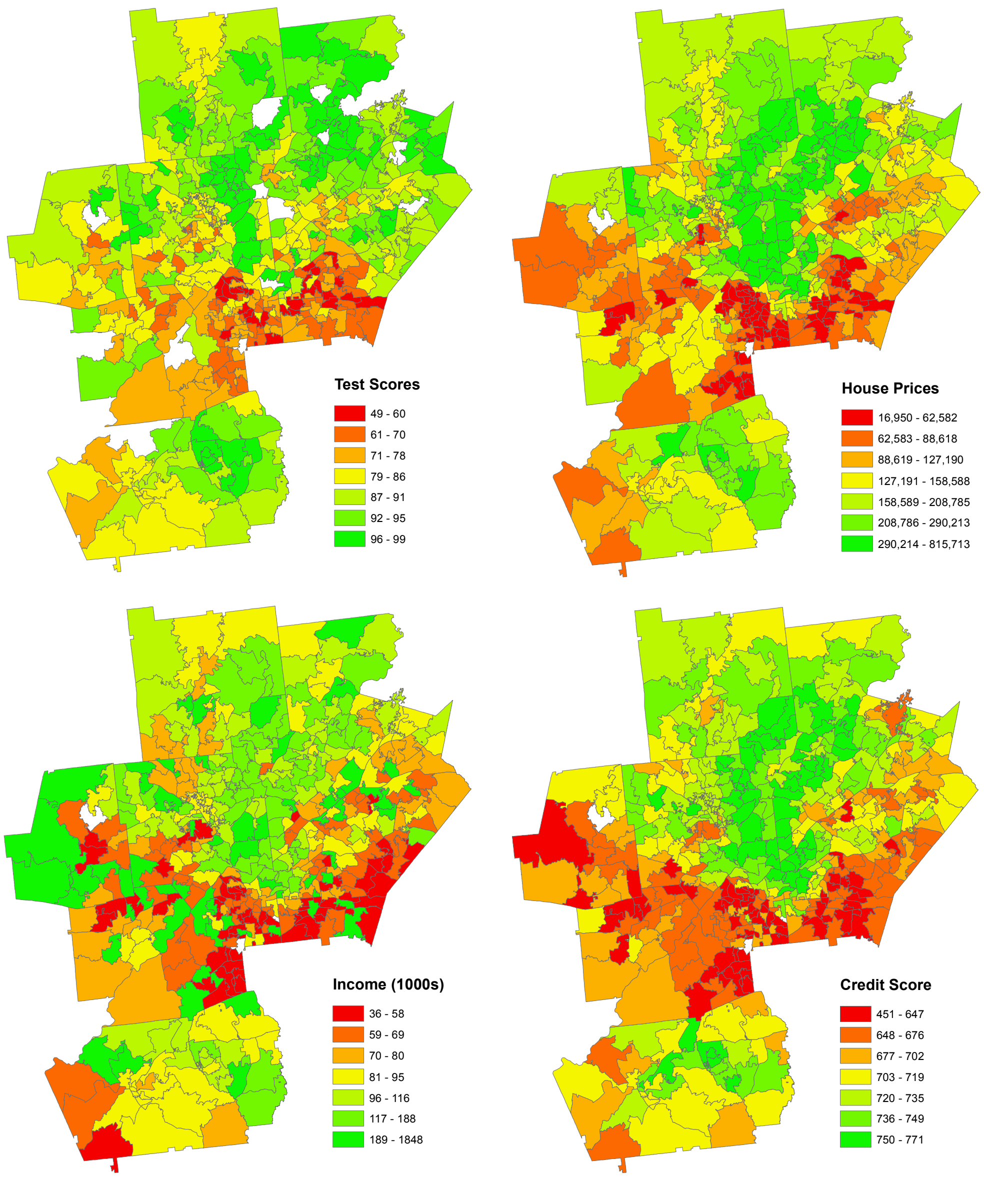
Walsh, Randy, "Endogenous open space amenities in a locational equilibrium," *Journal of Urban Economics*, 2007, *61* (2), 319 – 344.

Figure 1: Housing Transactions in Atlanta, 2011



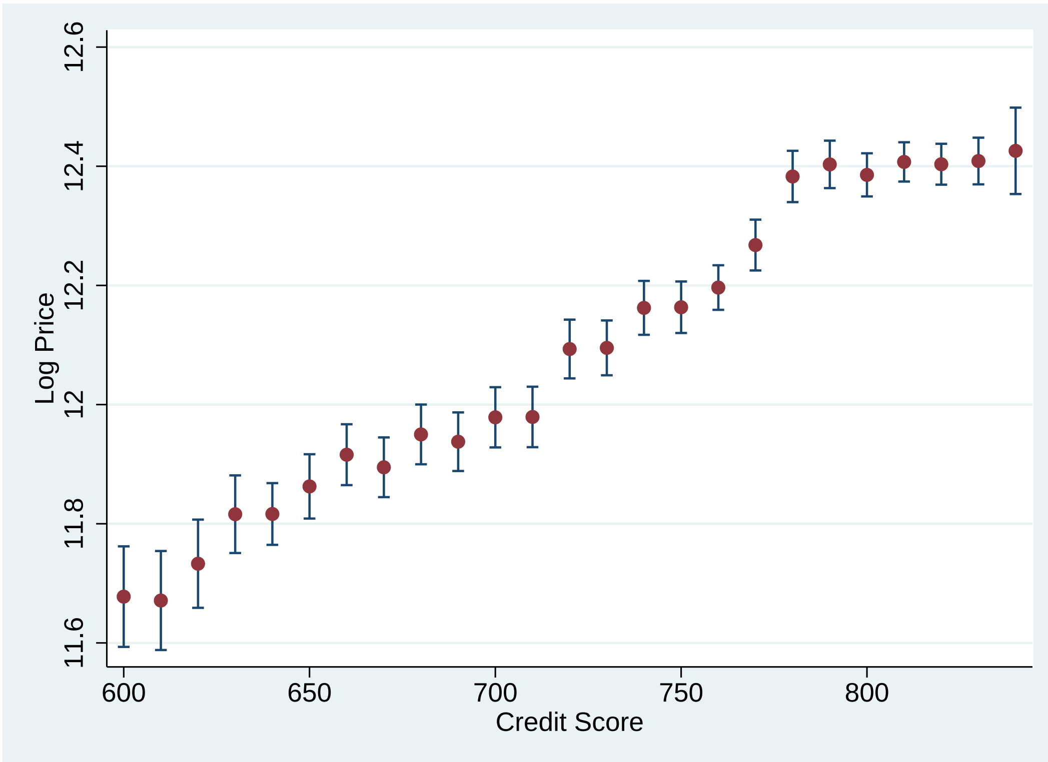
Note: This figure shows the Dataquick transactions in Atlanta in 2011 that were successfully matched to Experian credit reports. Each yellow dot is a housing transaction. The pink lines are elementary school attendance boundaries.

Figure 2: Test Scores, House Prices, Income, and Credit Scores in Atlanta, by Elementary School Attendance Zone



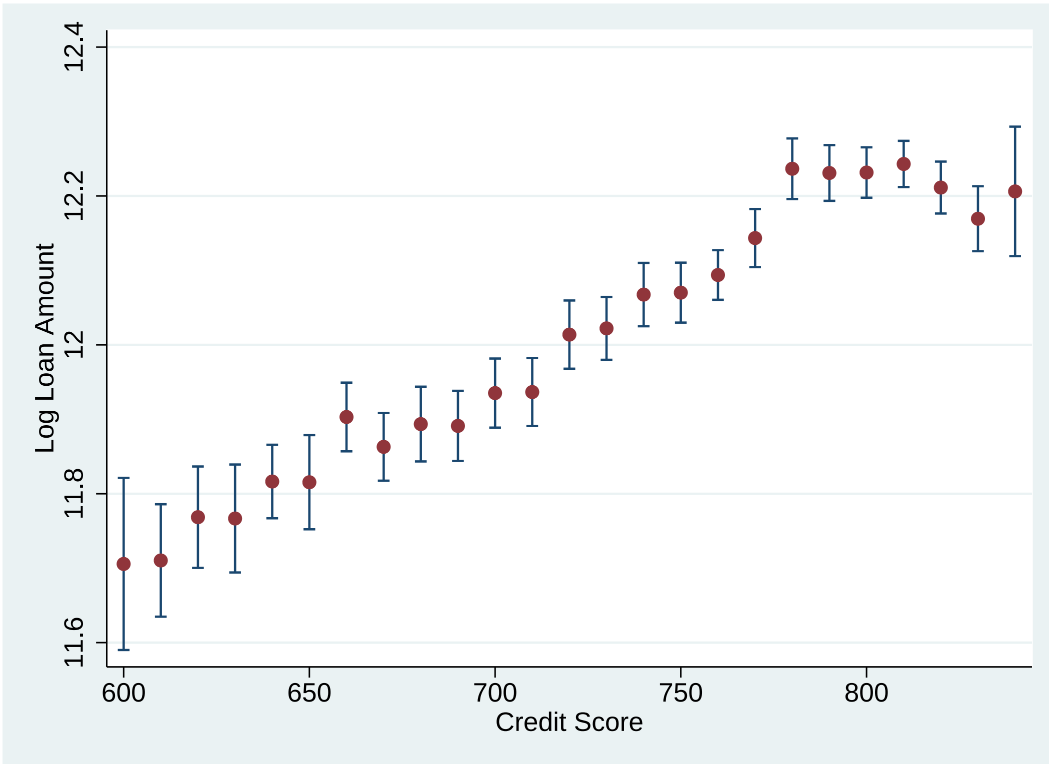
Note: The top left panel shows test scores for each school attendance zone in our data. The top right panel shows average house prices for the transactions in our data. The bottom left panel shows the average incomes of the buyers in our data and the bottom right panel shows the average credit scores for the buyers in our data. The figure illustrates the positive correlations between house prices, test scores, buyer incomes, and also credit scores.

Figure 3: Average House Price by Credit Score



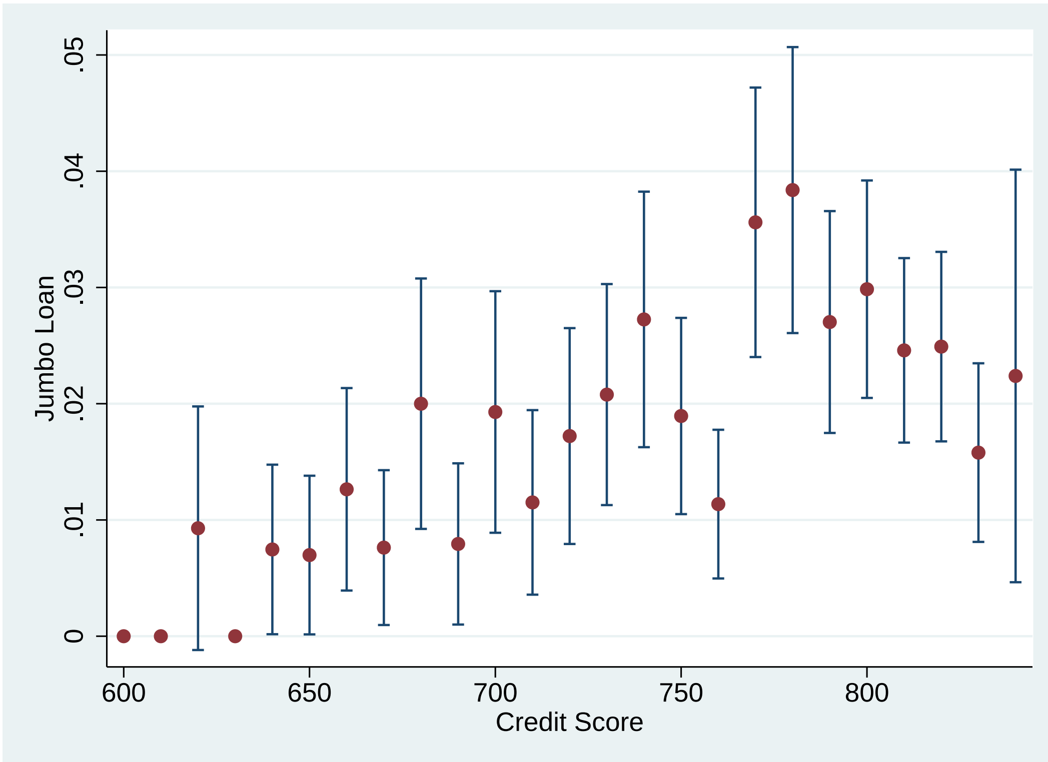
Note: This figure plots the average log house price against credit score using the buyers in the matched sample. Averages are computed in 10-point credit-score bins. Log price is increasing in credit scores, and there appears to be discontinuities at 720 and 770 credit score.

Figure 4: Average Loan Amount by Credit Score



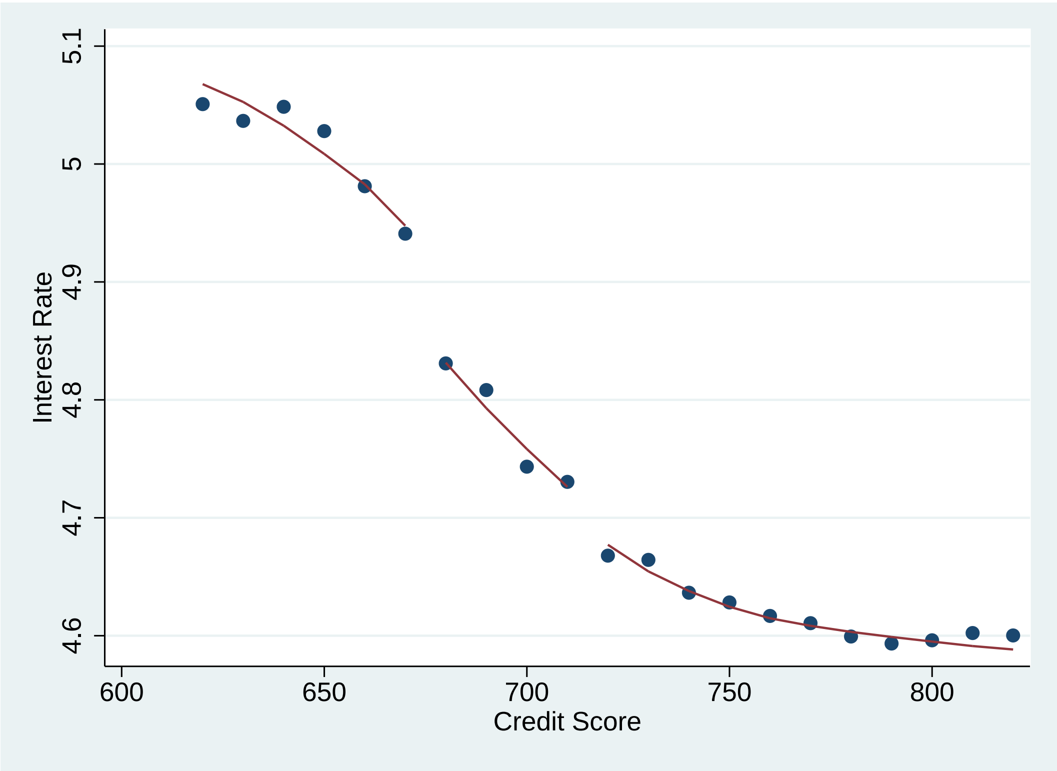
Note: This figure plots the average log loan amount against credit score using the buyers in the matched sample. Averages are computed in 10-point credit-score bins. Log loan amount is increasing in credit scores, and there appears to be discontinuities at 720 and 770 credit score.

Figure 5: Share of Jumbo Loans by Credit Score



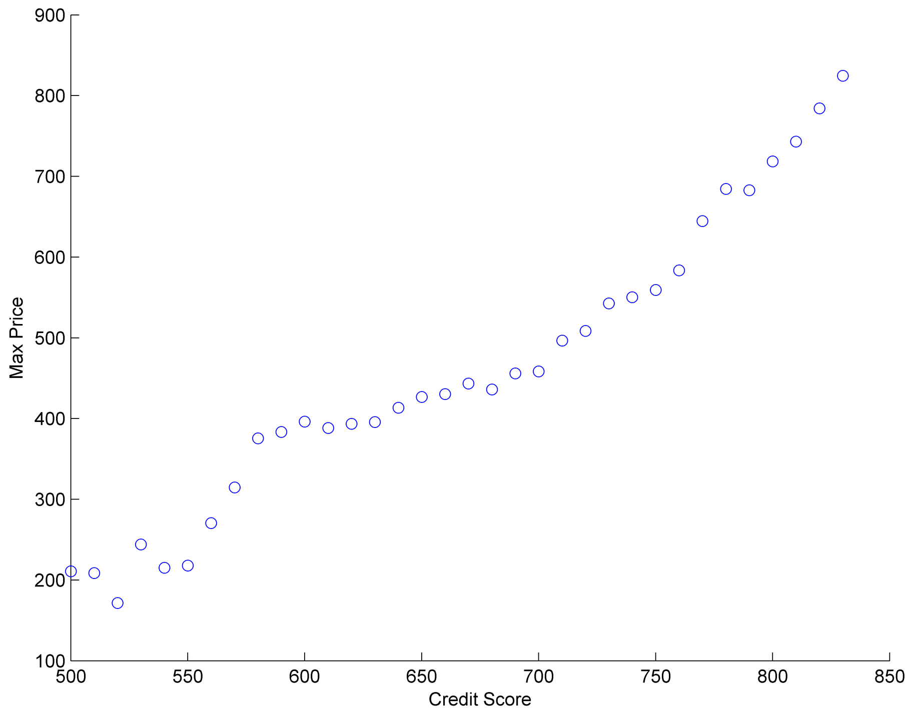
Note: This figure plots the share of jumbo loans against credit score using the buyers in the matched sample. Averages are computed in 10-point credit-score bins. A jumbo loan is defined as a loan with original balance greater than $417,000. There is a sharp discontinuity in the share of jumbo loans at 770 credit score.

Figure 6: Actual vs. Fitted Values of Interest Rates



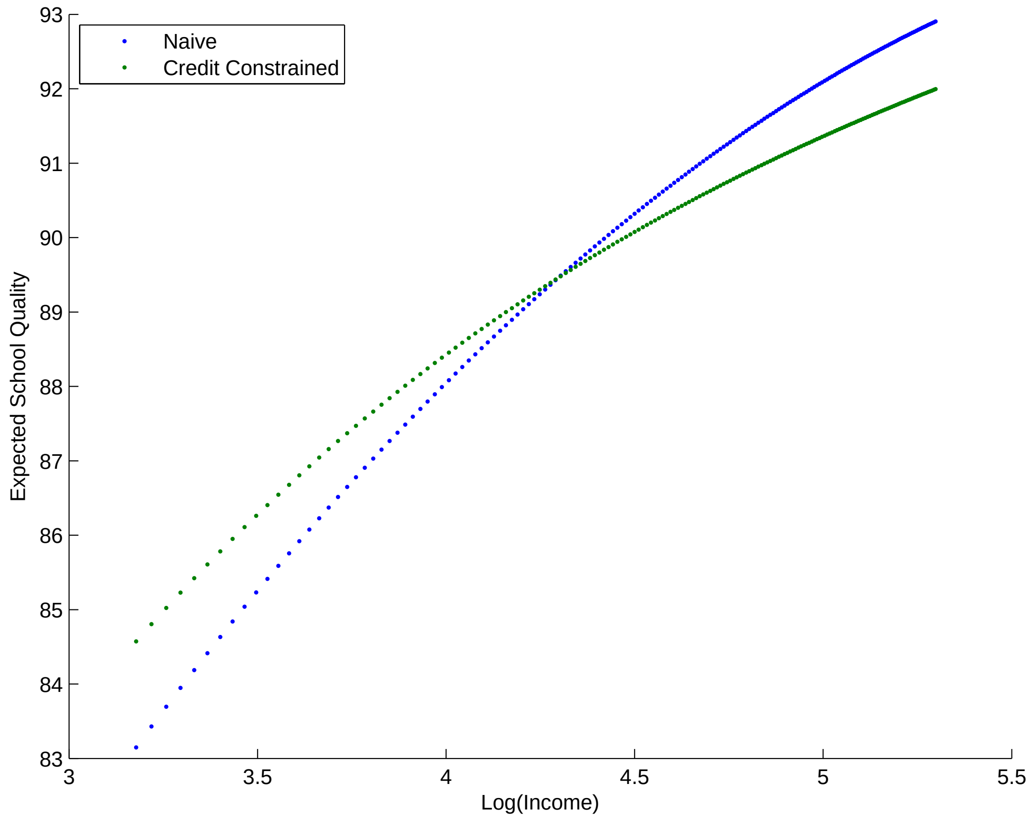
Note: This figure compares actual versus fitted interest rates using the Fannie Mae data and the interest rate model (12). The dots show average interest rate within 10-point credit score bins using the Fannie Mae data, while the line shows fitted interest rates using the estimation results reported in Table 8.

Figure 7: Average "House Price Frontier" by Credit Score



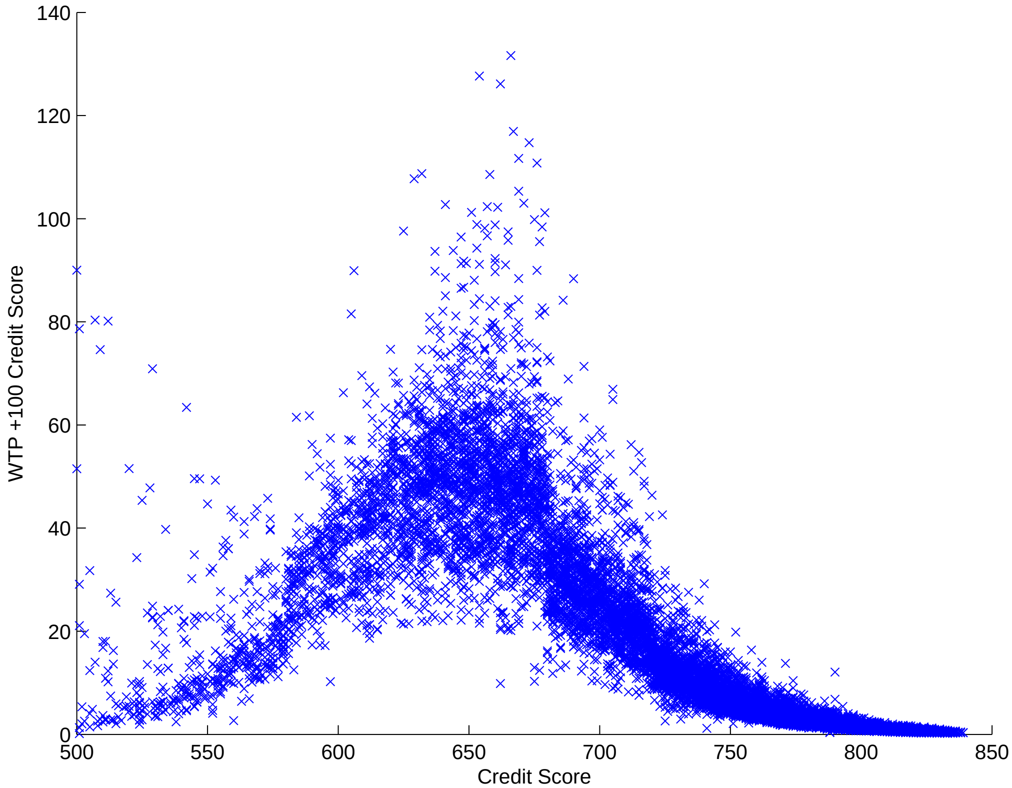
Note: This figure shows the "house price frontier", in $1,000s, averaged over 10-point credit score bins, computed using the formula described in (13).

Figure 8: Sorting on School Quality by Household Income



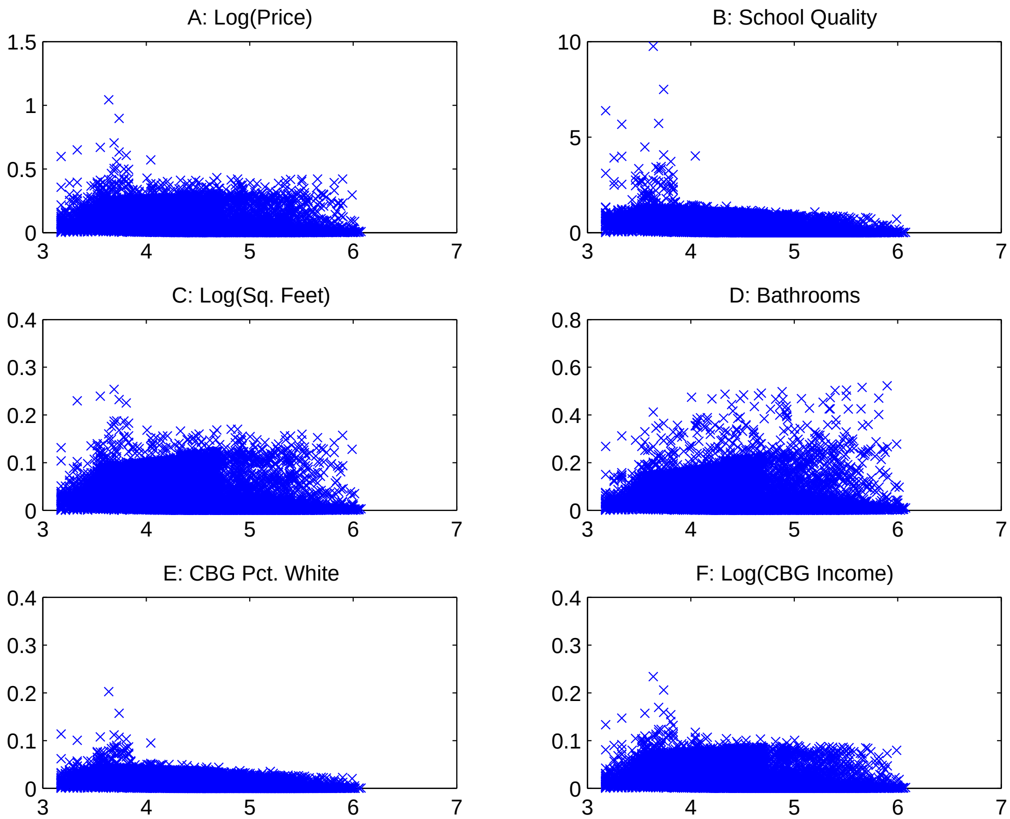
Note: This figure shows the expected choice of school quality by income for married, college-educated individuals under the age of 45, with 744 credit score and $15,000 downpayment (the population medians). The expected choice of school quality is shown for both the naive and credit constrained models. The implied sorting across income is stronger in the naive model than in the credit constrained model.

Figure 9: Willingness-to-Pay for +100 Credit Score



Note: This figure shows the willingness-to-pay for a 100 point credit score increase for all individuals in the data, reported in monthly dollars. In each counterfactual, only the credit score of one individual is increased. The minimum willingness-to-pay is 15 cents a month. The maximum is $131/month. The average is $17/month.

Figure 10: Expected Increase in Housing Attributes from a +100 Credit Score Increase



Note: This figure plots the expected increase in housing attributes that would result from a 100 point increase in credit score, for each individual in the data. In each counterfactual, only the credit score of one individual is increased. In each panel, the y-axis is the expected change in the attribute, and the x-axis is the individual's log annual income.

Table 1: Summary Statitsics for Matched Experian Data

	Mean	S.d.
Vantage Score	726	76
Vantage<680	0.275	0.446
680<=Vantage<720	0.138	0.345
Vantage>=720	0.588	0.492
Income Insight	97	63
Age	44	13
Male	0.631	0.483
White Collar	0.621	0.485
College Educated	0.533	0.499
Married	0.784	0.411
Matched Obs.	21,792	
Total Obs.	30,840	
Match Rate	70.7%	

Note: This table shows means and standard deviations for the Experian variables in our matched dataset. Vantage Score is a credit score measure that is equivalent in meaning to FICO. Income Insight is an estimate of buyer annual income, in $1,000s. Experian reports occupational data in 9 categories: management/business & financial operations; technical: computers/math & architect/engineering; professional: legal/education & health practitioner/tech/support; sales; office & administrative support; blue collar; farming/fish/forestry; other; retired. We consider the first five categories to be white-collar. College-educated indicates completion of bachelors' degree or higher.

Table 2: Summary Statitsics for Matched Housing and Neighborhood Data

	Mean	S.d.
Price	229,822	162,891
Test Score	87.4	9.6
Per. White	0.671	0.259
CBG Income	92,973	30,930
Sq. Feet	2,365	944
# Bedrooms	3.21	1.24
# Bathrooms	2.77	1.10
Total Obs.	21,792	
.5 miles	17,470	
.25 miles	11,136	
.2 miles	9,188	

Note: This table shows means and standard deviations for the housing and neighborhood attributes in our matched dataset. The statistics are taken from the full sample of matched observations. Test score is the performance of the local public elementary school on standardized tests. Percent white and CBG income are measured as the census-block-group level using 2010 Census data. All other variables are characteristics of the property itself, from Dataquick. "X miles" indicates the number of observations that are within X miles of the nearest school attendance boundary.

Table 3: Housing and Neighborhood Attributes: Unmatched and Matched

	Unmatched	Matched
Price	126,790	229,822
Test Score	82.4	87.4
Per. White	0.520	0.671
CBG Income	76,197	92,973
Sq. Feet	1,904	2,365
# Bedrooms	2.93	3.21
# Bathrooms	2.36	2.77
Total Obs.	9,048	21,792

Note: This table compares the mean of the housing and neighborhood attributes between the matched and unmatched samples. The houses in the matched sample are shown to be of higher quality than houses in the unmatched sample. This suggests that unmatched buyers are unmatched because they do not have credit reports, as opposed to being unmatched due to random errors like spelling inaccuracies.

Table 4: Regressions of House Price on Credit Score and Demographic Variables

	(1) Full Sample	(2) LTV>50%	(3) LTV<=50%	(4) Cash Buyers
Credit Score > 720	0.0385*	0.0352*	-0.0716	-0.0784
	(0.0202)	(0.0190)	(0.0692)	(0.0736)
Credit Score > 770	0.0513***	0.0496***	-0.000843	-0.00917
	(0.0198)	(0.0189)	(0.0645)	(0.0708)
Credit Score	-0.00138	-0.0375***	-0.0553	-0.0314
	(0.0128)	(0.0134)	(0.0339)	(0.0354)
Credit Score ^2	1.22e-05	5.97e-05***	8.66e-05*	4.97e-05
	(1.89e-05)	(1.96e-05)	(5.01e-05)	(5.25e-05)
Credit Score ^3	-1.02e-08	-3.10e-08***	-4.28e-08*	-2.42e-08
	(9.15e-09)	(9.49e-09)	(2.44e-08)	(2.56e-08)
Log(Income)	0.473***	0.494***	0.493***	0.465***
	(0.00908)	(0.00875)	(0.0272)	(0.0296)
Married	0.371***	0.315***	0.441***	0.428***
	(0.0105)	(0.0102)	(0.0302)	(0.0317)
College-Educated	0.249***	0.209***	0.363***	0.351***
	(0.00894)	(0.00850)	(0.0280)	(0.0301)
White-Collar	0.218***	0.185***	0.247***	0.199***
	(0.00920)	(0.00888)	(0.0276)	(0.0295)
Observations	17,376	14,118	3,258	2,796
R-squared	0.415	0.440	0.417	0.381

Note: This table reports the results from regression (1). Standard errors are given in parentheses. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively. In column 1, the full matched sample is used. In column 2, only buyers who purchased with loan-to-value ratio above 50% are included in the regression. In column 3, buyers who purchased with LTV below 50% are included. In column 4, only cash-buyers (i.e. non-borrowers) are included.

Table 5: F-Tests for the Joint Significance of Credit Score and Demographic Variables in House Prices

	(1) Full Sample	(2) LTV>50%	(3) LTV<=50%	(4) Cash Buyers
Credit Score				
F-statistic	2.94**	7.19***	0.63	0.38
p-value	0.0320	0.0001	0.5965	0.7687
Demographics				
F-statistic	1981.53***	1886.96***	306.78***	227.48***
p-value	0.0000	0.0000	0.0000	0.0000
Observations	17,376	14,118	3,258	2,796

Note: This table reports the results from two F-tests of the coefficients reported in Table 4. In the first test, the joint significance of the credit score variables are tested. In the second test, the joint significance of the demographic variables are tested. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively. In column 1, the full matched sample is used. In column 2, only buyers who purchased with loan-to-value ratio above 50% are included in the regression. In column 3, buyers who purchased with LTV below 50% are included. In column 4, only cash-buyers (i.e. non-borrowers) are included.

Table 6: Regressions of Housing Attributes on Credit Score and Demographic Variables

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	School Quality		Log Sq. Feet		# Bathrooms		CBG Pct. White	
	Full Sample	Cash Buyers	Full Sample	Cash Buyers	Full Sample	Cash Buyers	Full Sample	Cash Buyers
Credit Score > 720	0.0456**	0.0741	0.000870	-0.00570	0.0113	-0.0290	-0.0128*	0.00931
	(0.0216)	(0.0613)	(0.0104)	(0.0291)	(0.0325)	(0.0850)	(0.00765)	(0.0215)
Credit Score > 770	0.0210	0.0764	0.0125	0.0747***	-0.0290	-0.204**	-0.00862	0.00604
	(0.0215)	(0.0601)	(0.0103)	(0.0285)	(0.0323)	(0.0833)	(0.00761)	(0.0211)
Credit Score	0.0349**	0.00779	0.0114*	-0.00761	-0.00391	-0.0205	-0.0120**	0.00182
	(0.0136)	(0.0293)	(0.00655)	(0.0139)	(0.0205)	(0.0407)	(0.00483)	(0.0103)
Credit Score ^2	-4.97e-05**	-6.76e-06	-1.64e-05*	1.15e-05	6.82e-06	2.69e-05	1.88e-05***	-2.24e-06
	(2.01e-05)	(4.36e-05)	(9.65e-06)	(2.07e-05)	(3.02e-05)	(6.04e-05)	(7.11e-06)	(1.53e-05)
Credit Score ^3	2.33e-08**	1.01e-09	7.61e-09	-5.90e-09	-3.88e-09	-1.11e-08	-9.36e-09***	1.07e-09
	(9.75e-09)	(2.13e-08)	(4.68e-09)	(1.01e-08)	(1.47e-08)	(2.95e-08)	(3.45e-09)	(7.47e-09)
Log(Income)	-0.0332***	-0.0496*	0.115***	0.0826***	0.0500***	0.0730**	0.0448***	0.0482***
	(0.0103)	(0.0259)	(0.00488)	(0.0122)	(0.0155)	(0.0359)	(0.00364)	(0.00905)
Married	0.125***	0.0679**	0.170***	0.137***	-0.0784***	0.0171	0.0842***	0.0726***
	(0.0123)	(0.0287)	(0.00576)	(0.0134)	(0.0185)	(0.0398)	(0.00430)	(0.00997)
College-Educated	-0.0302***	-0.0406	0.00856*	-0.00912	0.0651***	0.106***	0.00482	0.00361
	(0.00969)	(0.0255)	(0.00466)	(0.0121)	(0.0146)	(0.0354)	(0.00343)	(0.00897)
White-Collar	0.0109	0.0236	0.0259***	0.00405	0.0566***	0.0990***	0.00259	-0.00401
	(0.00994)	(0.0249)	(0.00477)	(0.0118)	(0.0149)	(0.0345)	(0.00352)	(0.00874)
Observations	16,394	2,672	16,394	2,672	16,394	2,672	16,394	2,672
R-squared	0.459	0.522	0.508	0.508	0.384	0.412	0.379	0.450

Note: This table reports results from regression (1) with housing and neighborhood attributes on the LHS. Standard errors are given in parentheses. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively. In the odd-numbered columns, the full matched sample is used. In the even-numbered columns, only cash-buyers (i.e. non-borrowers) are included.

Table 7: F-Tests for the Joint Significance of Credit Score and Demographic Variables in Housing Attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	School Quality		Log Sq. Feet		# Bathrooms		CBG Pct. White	
	Full Sample	Cash Buyers	Full Sample	Cash Buyers	Full Sample	Cash Buyers	Full Sample	Cash Buyers
Credit Score								
F-statistic	3.04**	0.91	2.10*	2.38*	0.54	2.29*	2.11*	0.09
p-value	0.0276	0.4346	0.0978	0.0680	0.6528	0.0767	0.0970	0.9630
Demographics								
F-statistic	31.23***	3.07**	401.82***	39.45***	18.93***	7.27***	139.64***	21.23***
p-value	0.0000	0.0155	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000
Observations	16,394	2,672	16,394	2,672	16,394	2,672	16,394	2,672

Note: This table reports the results from two F-tests of the coefficients reported in Table 6. In the first test, the joint significance of the credit score variables are tested. In the second test, the joint significance of the demographic variables are tested. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively. In the odd-numbered columns, the full matched sample is used. In the even-numbered columns, only cash-buyers (i.e. non-borrowers) are included.

Table 8: Interest Rate Regressions

Parameter	Estimate	Std. Err.
τ_0	4.682***	(0.00997)
τ_1	0.426***	(0.0188)
τ_2	0.0364***	(0.00164)
τ_3	687.0***	(2.539)
τ_4	0.0778***	(0.00764)
τ_5	0.0210***	(0.00494)
τ_6	3.542***	(0.0858)
R^2	0.072	
N	259,587	

Note: This table reports estimation results for model (12) using the Fannie Mae data. All mortgages are 30-year fixed-rate loans, purchase loans, for owner-occupied single family homes or condos. Standard errors are given in parentheses. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively.

Table 9: Parameter Estimates for Naive Model

	(1) School Quality	(2) Log Sq. Feet	(3) # Baths	(4) CBG Pct. White	(5) Log CBG Income	(6) Cost
Constant	0.03812	25.57***	4.604***	4.744***	3.405***	3.477***
	(0.02569)	(0.4253)	(0.1613)	(1.133)	(0.6101)	(0.01296)
Log Income	0.02504***	0.5077***	-0.09216***	0.8371***	0.7135***	-0.144***
	(0.0007041)	(0.02212)	(0.008017)	(0.02883)	(0.02162)	(0.002758)
Middle Age	-0.003957	0.5376***	-0.04574*	-0.06666	-0.001955	0.03046***
	(0.002813)	(0.0796)	(0.02511)	(0.1067)	(0.08171)	(0.003929)
Old Age	-0.004854	-0.02745	-0.2703***	0.1608	-0.1549	0.05504***
	(0.005813)	(0.162)	(0.05123)	(0.2091)	(0.1608)	(0.00877)
College Educated	-0.007961***	-0.2812***	0.01645	0.1783*	0.5848***	-0.07139***
	(0.002629)	(0.07462)	(0.02455)	(0.09987)	(0.0769)	(0.003956)
Married	0.03198***	1.577***	-0.0323	0.5732***	1.071***	-0.03537***
	(0.00293)	(0.09295)	(0.03458)	(0.1199)	(0.09079)	(0.005947)
Log Likelihood	-116696.88					
N	12,817					

Note: This table reports the parameter estimates for the model given in Sections 4 and 5.1. "Cost" is measured as the annualized cost of the house, in thousands of dollars. In the naive model, there is no interest heterogeneity and all houses are in each individual's choice set. The interest rate used to annualize house prices is the average interest rate across all individuals, as computed by (12). Standard errors are given in parentheses. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively.

Table 10: Parameter Estimates for Credit Constrained Model

	(1) School Quality	(2) Log Sq. Feet	(3) # Baths	(4) CBG Pct. White	(5) Log CBG Income	(6) Cost
Constant	0.00549	20.66***	3.418***	1.131	3.655***	2.558***
	(0.01971)	(0.3264)	(0.1238)	(0.8695)	(0.4682)	(0.01562)
Log Income	0.02411***	0.08059***	-0.03469***	1.173***	0.2909***	-0.08853***
	(0.0007046)	(0.02228)	(0.008029)	(0.0289)	(0.02163)	(0.003393)
Middle Age	-0.0007509	0.5057***	0.001247	-0.1268	-0.07187	0.05863***
	(0.002813)	(0.08025)	(0.02512)	(0.107)	(0.08195)	(0.004701)
Old Age	-0.002348	0.1556	-0.1259**	0.3268	-0.4802***	0.1419***
	(0.005829)	(0.1633)	(0.05178)	(0.2112)	(0.163)	(0.009787)
College Educated	-0.002873	-0.2829***	0.03091	-0.118	0.4037***	-0.07752***
	(0.00263)	(0.07535)	(0.02457)	(0.1003)	(0.07713)	(0.004645)
Married	0.02631***	2.239***	-0.1382***	1.041***	1.106***	-0.03587***
	(0.002934)	(0.09393)	(0.03463)	(0.1204)	(0.09081)	(0.006586)
Log Likelihood	-113209.58					
N	12,817					

Note: This table reports the parameter estimates for the model given in Sections 4 and 5.2. "Cost" is measured as the annualized cost of the house, in thousands of dollars. In the credit constrained model, there is interest heterogeneity as described in Section 4.1, and the choice sets are restricted as described in Section 4.2. Standard errors are given in parentheses. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively.

Table 11: Parameter Estimates (Credit Constrained − Naive)

	(1) School Quality	(2) Log Sq. Feet	(3) # Baths	(4) CBG Pct. White	(5) Log CBG Income	(6) Cost
Constant	-0.03263	-4.912***	-1.186***	-3.612**	0.2498	-0.9191***
	(0.03238)	(0.5361)	(0.2033)	(1.428)	(0.7691)	(0.02029)
Log Income	-0.0009337	-0.4271***	0.05747***	0.336***	-0.4226***	0.05543***
	(0.0009961)	(0.03139)	(0.01135)	(0.04082)	(0.03058)	(0.004373)
Middle Age	0.003206	-0.03195	0.04699	-0.06012	-0.06992	0.02817***
	(0.003978)	(0.113)	(0.03551)	(0.1512)	(0.1157)	(0.006127)
Old Age	0.002506	0.183	0.1444**	0.166	-0.3253	0.0869***
	(0.008232)	(0.2301)	(0.07284)	(0.2971)	(0.229)	(0.01314)
College Educated	0.005088	-0.001709	0.01446	-0.2964**	-0.1811*	-0.006136
	(0.003719)	(0.106)	(0.03474)	(0.1415)	(0.1089)	(0.006101)
Married	-0.005673	0.6617***	-0.1059**	0.4676***	0.03505	-0.0004959
	(0.004147)	(0.1321)	(0.04894)	(0.1699)	(0.1284)	(0.008874)
Log Likelihood	+3487.30					
N	12,817					

Note: This table reports the differences in parameter estimates between the credit constrained model and the naive model. Standard errors of the differences are given in parentheses. ***, **, and * indicate p-values of 0.01, 0.05, and 0.1 respectively.

Table 12: Willingness-to-Pay Estimates, by Model and Subpopulation

	(1) +10 School Quality	(2) +10% Sq. Feet	(3) +1 Bathrooms	(4) +10 CBG Pct. White	(5) +10% CBG Income
Panel A: Full Sample					
Naïve	50.49	87.14	124.1	26.95	23.2
Credit Constrained	51.62	89.7	124.2	27.78	23.61
Pct. Diff.	2.235	2.939	0.1209	3.081	1.764
Panel B: $55k income, college educated, married					
Naïve	48.47	86.19	125.8	26.4	23.62
Credit Constrained	50.05	91.43	126.4	26.93	25.24
Pct. Diff.	3.261	6.079	0.4789	2.025	6.832
Panel C: $116k income, college educated, married					
Naïve	56.21	90.82	128.7	29.39	26.22
Credit Constrained	59.09	94.67	129.5	31.41	26.96
Pct. Diff.	5.134	4.239	0.5943	6.887	2.803

Note: This table reports the willingness-to-pay estimates implied by the coefficient estimates in Tables 9 and 10. Willingness-to-pay is computed as the marginal utility of an attribute divided by the marginal utility of money. WTP is reported in dollars-per-month.

Table 13: Expected Increase in Housing Attributes from a +100 Credit Score Increase (Summary)

	Expected Increase in Attribute		
Housing Attribute	Mean	Max	Corr. with Income
Log(Price)	0.07935	1.043	-0.2953
School Quality	0.3295	9.754	-0.3931
Log(Sq. Feet)	0.02812	0.2537	-0.2449
Bathrooms	0.051	0.5224	-0.1316
CBG Pct. White	0.01104	0.2026	-0.403
Log(CBG Income)	0.02045	0.2339	-0.3001

Note: This table summarizes the expected increase in housing attributes that would result from a 100 point increase in credit score, for each individual in the data. In each counterfactual, only the credit score of one individual is increased.

Table 14: The Effect of Credit Constraints on Residential Sorting

	With C.C. (Baseline)	No C.C. (Counterfactual)	Pct. Explained by C.C.
Panel A: Standard deviation in amenities consumed			
School Quality	4.205	3.574	15.0%
Log(Sq. Feet)	0.2246	0.1744	22.4%
Bathrooms	0.446	0.3256	27.0%
CBG Pct. White	0.1178	0.09421	20.0%
Log(CBG Income)	0.1906	0.1569	17.7%
Panel B: Correlation of amenities consumed with income			
School Quality	0.6447	0.6054	6.1%
Log(Sq. Feet)	0.6474	0.5699	12.0%
Bathrooms	0.6567	0.6128	6.7%
CBG Pct. White	0.6793	0.6475	4.7%
Log(CBG Income)	0.6795	0.6173	9.2%

Note: This table reports the standard deviations in amenities consumed and the correlation between amenities consumed with income, as simulated under two models. The first model is the baseline discrete choice model with credit constraints as discussed in Section 4. In the second model, every individual is assumed to face a 5.07% mortgage interest rate, regardless of credit score or borrowing amount, and all houses are assumed to be in all individuals' choice sets.

A Proof of the contraction mapping

We begin with a restatement of the theorem given in Berry et al. (1995):

Theorem. *Consider the metric space* $\left(\mathbb{R}^K, d\right)$ *with* $d(x,y) = \|x - y\|$ *(where* $\|\cdot\|$ *is the sup-norm). Let* $f : \mathbb{R}^K \to \mathbb{R}^K$ *have the properties:*

1. $\forall x \in \mathbb{R}^K$, $f(x)$ *is continuously differentiable, with,* $\forall j, k$,

$$\partial f_j(x) / \partial x_k \geq 0$$

and

$$\sum_{k=1}^{K} \partial f_j(x) / \partial x_k < 1$$

2. $\min_j \inf_x f(x) \equiv \underline{x} > -\infty$

3. *There is a value* $\bar{x}$ *with the property that if for any* j, $x_j \geq \bar{x}$, *then for some* k *(not necessarily equal to* j*),* $f_k(x) < x_k$.

Then there is a unique fixed point, x_0, *to* f *in* $\mathbb{R}^K$. *Further, let the set* $X = [\underline{x}, \bar{x}]^K$, *and define the truncated function,* $\hat{f} : X \to X$, *as* $\hat{f}_j(x) = \min\{f_j(x), \bar{x}\}$. *Then,* $\hat{f}(x)$ *is a contraction mapping of modulus less than one on* X.

Our function of interest is given in (11):

$$f_h(\delta) = \delta_h + \log s_h - \log \hat{s}_h(\delta)$$

where

$$\begin{aligned} \hat{s}_h(\delta) &= \frac{1}{N} \sum_{i=1}^{N} \sum_{q=1}^{H} \phi_{iq} P_{ih}^q \\ P_{ih}^q &= \begin{cases} \frac{\exp(\lambda_{ih} + \delta_h)}{\sum_{j=1}^{q} \exp(\lambda_{ij} + \delta_j)} & \text{if } h \leq q \\ 0 & \text{if } h > q \end{cases} \end{aligned}$$

For convenience in writing the proof, we let houses h be indexed from 0 to H, so there are $H+1$ houses, and $h=0$ is the house whose "delta" is normalized to zero. Because "delta" for $h=0$ is normalized, f is here an H-dimensional function of the H-dimensional vector δ. A useful derivative will be:

$$\frac{\partial P_{ih}^{q}}{\partial \delta_k} = \begin{cases} P_{ih}^{q}\left(1-P_{ih}^{q}\right) & \text{if } h=k \\ -P_{ih}^{q}P_{ik}^{q} & \text{otherwise} \end{cases}$$

To show the first condition of the theorem, note:

$$\begin{aligned} \frac{\partial f_h}{\partial \delta_h} &= 1-\frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}\partial P_{ih}^{q}/\partial\delta_h}{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}} \\ &= 1-\frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}\left(1-P_{ih}^{q}\right)}{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}} \geq 0 \end{aligned}$$

$$\begin{aligned} \frac{\partial f_h}{\partial \delta_k} &= -\frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}\partial P_{ih}^{q}/\partial\delta_k}{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}} \\ &= \frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}P_{ik}^{q}}{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}} \geq 0 \end{aligned}$$

and:

$$\begin{aligned} \sum_{k=1}^{H}\frac{\partial f_h}{\partial \delta_k} &= 1-\frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}\left(1-P_{ih}^{q}\right)}{\sum_{i=1}^{N}\sum_{q=1}^{Q}\phi_{iq}P_{ih}^{q}}+\frac{\sum_{k\neq h}\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}P_{ik}^{q}}{\sum_{i=1}^{N}\sum_{q=1}^{Q}\phi_{iq}P_{ih}^{q}} \\ &= 1-\frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}\left[P_{ih}^{q}\left(1-P_{ih}^{q}\right)-\sum_{k\neq h}P_{ih}^{q}P_{ik}^{q}\right]}{\sum_{i=1}^{N}\sum_{q=1}^{Q}\phi_{iq}P_{ih}^{q}} \\ &= 1-\frac{\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}\left[1-P_{ih}^{q}-\sum_{k\neq h}P_{ik}^{q}\right]}{\sum_{i=1}^{N}\sum_{q=1}^{Q}\phi_{iq}P_{ih}^{q}} \\ &= 1-\frac{\sum_{i=1}^{N}\sum_{q=1}^{Q}\phi_{iq}P_{ih}^{q}P_{i0}^{q}}{\sum_{i=1}^{N}\sum_{q=1}^{Q}\phi_{iq}P_{ih}^{q}} < 1 \end{aligned}$$

The second condition of the theorem is that $\min_k \inf_\delta f_k(\delta) > -\infty$, or, f

has a lower bound. To show that this is true, we first note:

$$\begin{aligned}
f_h(\delta) &= \delta_h - \log\left(\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}P_{ih}^{q}\right) \\
&= \delta_h - \log\left(\sum_{i=1}^{N}\sum_{q=h}^{H}\phi_{iq}\frac{\exp(\lambda_{ih}+\delta_h)}{\exp(\lambda_{i0})+\sum_{j=1}^{q}\exp(\lambda_{ij}+\delta_j)}\right) \\
&= -\log\left(\sum_{i=1}^{N}\sum_{q=h}^{H}\phi_{iq}\frac{\exp(\lambda_{ih})}{\exp(\lambda_{i0})+\sum_{j=1}^{q}\exp(\lambda_{ij}+\delta_j)}\right)
\end{aligned}$$

We already established that f_h is increasing in each δ_k. This equation shows that:

$$\lim_{\delta\to-\infty} f_h(\delta) = -\log\left(\sum_{i=1}^{N}\sum_{q=1}^{H}\phi_{iq}\frac{\exp(\lambda_{ih})}{\exp(\lambda_{i0})}\right)$$

so f_h is indeed bounded below.

The final condition of the theorem is that there exists a value $\bar{\delta}$ such that if any $\delta_k \geq \bar{\delta}$, then for some ℓ (not necesssarily equal to k), $f_\ell(\delta) < \delta_\ell$. Following Berry (1994), define $\bar{\delta}_k$ as the δ_k which sets $\hat{s}_0(\delta)$ to $1/N$ when all the other δ_ℓ's for $\ell \neq k$ are $-\infty$. In words, it is the δ_k which would set the predicted market share of the outside option to its data equivalent, when all other options besides k are considered unchooseable. Then, define $\bar{\delta} = \max_k \bar{\delta}_k$. If δ is such that $\delta_k \geq \bar{\delta}$ for some k, then $\hat{s}_0(\delta) < 1/N$, and so:

$$\sum_{h=1}^{H}\hat{s}_h(\delta) > 1 - \frac{1}{N} = \frac{N-1}{N}$$

which implies that $\hat{s}_\ell(\delta) > 1/N$ for some ℓ. Therefore:

$$f_\ell(\delta) = \delta_\ell + \log s_\ell - \log\hat{s}_\ell(\delta) < \delta_\ell$$

This satisfies the final condition. Therefore, equation (11) has a unique fixed point and the equation:

$$\bar{f}_h(\delta) = \min\left\{\delta_h + \log s_h - \log\hat{s}_h(\delta), \bar{\delta}\right\}$$

is a contraction mapping also with a unique fixed point. Moreover, the unique fixed point of $\bar{f}$ is also the unique fixed point of f.