Structural Demand Estimation with Borrowing Constraints*

Amine Ouazad† Romain Rancière‡

October 30, 2015

Abstract

Structural models of location choice use observed demand to estimate household preferences. However, household demand may be partly determined by borrowing constraints, limiting households' choice set. Credit availability differs across locations, households, and years. We put forward a model of neighborhood choice where mortgage approval rates determine households' choice set. Using household-level data, geocoded transactions, and mortgage applications for the San Francisco Bay area, we find that including borrowing constraints leads to higher estimated preferences for better performing schools and majority-white neighborhoods. General equilibrium estimates of the relaxation of lending standards provide two out-of-sample predictions: between 2000 and 2006, (i) a compression of the price distribution and (ii) a decline in black exposure to Whites. Both predictions are supported by empirical observation.

Keywords: Location choice, industrial organization, urban economics, credit constraints, differentiated products, segregation.

JELs: R21, R23, G21

*We thank Patrick Bayer, Jenna Bednar, Gilles Duranton, Ambika Gandhi, Maria Guadalupe, Timothy Van Zandt, and the INSEAD and University of Southern California seminar series for fruitful comments. Marton Varga provided research assistance. The authors acknowledge financial support from INSEAD, the International Monetary Fund, the Chaire Banque de France.

†INSEAD, Boulevard de Constance, Fontainebleau. *amine.ouazad@insead.edu*.

‡Paris School of Economics, CEPR and IMF

1 Introduction

The ability to secure mortgage credit is a key determinant of households' decision to purchase housing. Changes in lending standards may thus potentially shape the structure of cities. This paper proposes and estimates a novel structural model of housing demand under borrowing constraints, and estimates it to understand the role played by credit in shaping housing demand. The model allows general equilibrium analysis of the effects of changes in lending standards on the distribution of prices and on spatial segregation.

Structurally estimated models of consumer demand for differentiated goods have been widely used to characterize the market for cars (Berry, Levinsohn & Pakes 1995), cereals (Nevo 2000), newspapers (Fan 2013), and housing (Bayer & Timmins 2007, Bayer, Ferreira & McMillan 2007). With limited data requirements, such models estimate the full set of own-price and cross-price demand elasticities, as well as income elasticities. However, existing models of consumer demand for differentiated goods, pioneered by Berry et al. (1995), do not explicitly account for credit constraints. Credit constraints could be empirically important in housing markets – more than 80% of purchases involve mortgage financing.[1] Applied to housing, these models thus face the risk of attributing to household preferences features of household demand that in fact reflect households' borrowing constraints.

This paper is, to the best of our knowledge, the first to extend the framework of Berry et al. (1995) to account for the effect of endogenous borrowing constraints on household demand. Using data on mortgage credit applications, transaction prices, as well as individual and neighborhood characteristics, the model estimates households' preferences and willingness to pay for housing amenities in the San Francisco Bay area between 1990 and 2010.

The key empirical findings of the paper can be summarized as follows. First, borrowing constraints significantly impact the estimation of preferences for housing and neighborhood characteristics. When mortgage origination constraints are taken into account, households exhibit a higher willingness to pay for schools with higher test scores, as well as for mostly white neighborhoods. Second, the model with borrowing constraints allows a novel decomposition of the own-price demand elasticity into two distinct parts: (i) a fraction of elasticity that reflects the impact of a price

[1] 81% in 2013 according to data provided by RealtyTrac.

increase on utility, and (ii) a fraction that reflects the impact of a price increase on mortgage approval rates. Indeed, a higher loan-to-income ratio reduces the probability of approval. Third, we run a general equilibrium analysis of the impact of a change in lending standards of the magnitude experienced between 2000 and 2006. In doing so, we show that our estimated model explains key features of the actual distribution of house price changes across neighborhoods, as well as observed patterns of changes in racial exposures, including the mobility of white households to majority-white neighborhoods.

This paper's model aims at retaining the tractability and identification properties of the Berry et al. (1995) class of models while incorporating estimated borrowing constraints. The key feature of the model is that the likelihood of approval for mortgage applications in each neighborhood directly affects the choice set of potential homebuyers – a typically substantially smaller set of neighborhoods than the entire metropolitan area. Mortgage originators approve or deny borrowers' applications based on a screening process that combines information on loan amount and leverage (loan-to-income and loan-to-value ratios) with borrower creditworthiness and other borrower characteristics, property value, and neighborhood prospects.

Absent mortgage credit denial risk, households' neighborhood demand results from a discrete choice model, in which housing choices reflect neighborhood and housing characteristics, as well as the interaction between neighborhood and housing characteristics. In that case, the intuition of the standard Alonso et al. (1964) holds: households' preferences for neighborhood amenities are fully reflected in price differences. This may no longer be the case when borrowing constraints, resulting from the risk of being denied mortgage credit, are at play. In this case, households face different probabilities of access to mortgage credit for different neighborhoods – based on banks' lending standards. The differential availability of mortgage credit may prevent households from fully arbitraging across locations; and prices may not reflect the value of local amenities.

The probability of approval for a loan application is estimated using Home Mortgage Disclosure Act (HMDA) data, which includes both the income and race of the applicant, and is matched to longitudinal data on neighborhood time-varying characteristics – including geocoded transaction prices. In estimating approval probabilities, we exploit the very unique characteristics of HMDA data which is to provide information on the universe of *mortgage applications* and their associated *approval decisions.*[2] The coefficients of the probability of approval regression describe how banks' lending standards convert individual, property, and neighborhood characteristics into a probability of mortgage loan approval. Lending standards constrain households' neighborhood choice set, likely affecting household demand. By comparing the results of the same model estimated with and without borrowing constraints, we test whether borrowing constraints bias estimated household preferences: preferences for neighbors' race, income, education, for houses' year of construction, number of rooms, and for school test scores and distance to the central business district.

Introducing borrowing constraints raises at least two estimation challenges. In the model without borrowing constraints, Berry et al. (1995) have shown how observed demand maps one-to-one with neighborhood utilities.[3] This paper extends this key result to the case in which each household demand is conditional on the set of neighborhoods for which households' mortgage applications can be approved with some probability. Household preferences for amenities are then estimated using a simulated method of moments estimator.[4] The second issue is in the estimation of mortgage approval probabilities, which is subject to potential endogeneity bias due to unobservable borrowers' characteristics. We estimate the impact of borrower and mortgage characteristics using a set of instruments combining information on banks' balance sheet liquidity, at the national level, with information on the location of their branches across neighborhoods. Bank liquidity, derived from national balance sheets, predicts local loan characteristics, and is unlikely to be confounded by local unobservable demand factors. Finally, house prices are typically endogenous with respect to local unobservables, an issue shared with prior consumer choice models; this leads to an upward bias in the estimated impact of prices on neighborhood utility. We use the set of characteristics of adjacent neighborhoods as instruments for neighborhood price, in a similar way as Bayer & Timmins (2007).

The model yields estimates of household preferences for 4,416 neighborhoods (i.e. census block-groups) in the San Francisco Bay area. In the model's estimation, household preferences vary along seven broad dimensions: house price, neighbors' observables, housing quality, distance to the central business district, and school test scores as measured by California's Academic Performance

[2] Note that other datasets (e.g. loan performance data) might include additional informations on the loan and on the borrower but only for *originated loans* not for applications, and they do not include approval decisions.

[3] The contraction mapping is shown to yield a unique vector of utilities, up to a constant, for every observed vector of aggregate product demand.

[4] The total number of potential choice sets leads to our use of Simulated GMM: 2 to the power of the number of neighborhoods is a very large number, leading to computationally intractable estimation when using non-simulated demand.

Index. The model also estimates household preference heterogeneity using census data for the 1% Census' 120,029 households: preferences for neighborhoods vary according to the household's race, income, and unobserved heterogeneity. The model with borrowing credit constraints yields stronger estimated preferences for mostly-white neighborhoods, and for neighborhoods with high-performing public schools. By contrast the model without borrowing constraints tends to underestimate households' willingness to pay for such amenities; as households are constrained by low mortgage approval probabilities in such locations.

The model with borrowing constraints gives a higher sensitivity of demand to a change in house prices. Indeed, the model with borrowing constraints features two kinds of demand elasticities: conditional and total (own-price) demand elasticity. Conditional demand elasticity measures the effect of a price change on demand assuming that mortgage approval probabilities and the resulting neighborhood choice set are left unchanged. In addition to this effect, total demand elasticity takes also into account the effect of price change on mortgage approval-based choice set probabilities.

Our estimation results suggest that total demand elasticity can be 20 to 40 percent larger, in absolute value, as compared to the usual conditional demand elasticity and exhibit a substantially larger degree of heterogeneity across households, reflecting differences in the intensity of the borrowing constraints.

Finally, comparative statics are possible as city equilibrium is globally unique under specific conditions, and locally unique under general conditions. We thus estimate how lending standard changes affect house prices and spatial segregation in general equilibrium. Our first comparative statics result shows that a relaxation in lending standards, as observed in the Bay area between 2000 and 2006, leads to a compression of the price distribution. Both in the model and in the data, neighborhoods that were less expensive in 2000 saw a larger increase between 2000 and 2006 than initially more expensive neighborhoods.[5] We also analyze the effect of the same change in lending standards on spatial segregation. Very much like in the data (Ouazad & Rancière Forthcoming), the model predicts a reduction in the exposure of Blacks to Whites and an increase in black exposure to Hispanics and Asians.

We test the robustness of our comparative results to the introduction of a neighborhood-specific

[5] A similar cross-sectional dispersion in house price changes has been documented for San Diego by Landvoigt, Piazzesi & Schneider (2015).

supply elasticity (constructed using satellite data from the U.S. Geological Survey on land development combined with local measures of land slope and ruggedness), and to the introduction of tenure choice (rental vs. homeownership) in each neighborhood. In both cases, the robustness tests yield results that are similar to our baseline findings.

This paper makes several contributions to the literature. The paper introduces borrowing constraints in the workhorse model of the consumer choice literature developed in Berry, Levinsohn & Pakes (1995) and Petrin (2002) for consumer products, in Bayer & Timmins (2007) and Bayer, Ferreira & McMillan (2007) for housing.[6] The paper models borrowing constraints as households' choice sets. By focusing on the effect of borrowing constraints in reducing the choice set, our paper relates to already existing literature on incomplete product availability (Conlon & Mortimer 2013) or limited information on products (Goeree 2008). Our approach, however, differs from that literature in two important ways. First, the reduction of the choice set in our model is due to the endogenous decision of banks to approve mortgage loan applications. Second, the reduction of the choice set in our model comes from the (housing) demand side rather than for the supply side.

The general equilibrium comparative statics analysis is another methodological contribution of the paper. Such analysis is particularly useful in the case of housing (as opposed to consumer products) because housing supply elasticity is very far from perfectly elastic, and therefore demand shocks typically lead to substantial changes in the price distribution. By looking at the impact of changes in lending standards on the equilibrium distribution of prices across neighborhoods, this paper contributes to the literature linking changes in credit conditions to changes in house prices (Favara & Imbs 2015, Adelino, Schoar & Severino 2012, Glaeser, Gottlieb & Gyourko 2012, Landvoigt et al. 2015). Our model accounts for the observed within city compression in the distribution of house prices following a relaxation of lending standards. Landvoigt, Piazzesi & Schneider (2015), using a calibrated assignment model, shows how the relaxation of collateral constraints replicates this empirical finding. However, our paper is, to our knowledge, the first in which this pattern emerges, out-of-sample, within a structurally *estimated* model of housing with borrowing constraints featuring heterogeneous households, heterogeneous neighborhoods, and social interactions. The construction of a novel set of instruments for predicting local loan characteristics is a side contribution to this literature.

[6] See Holmes & Sieg (2015) for a comprehensive survey of structural estimation in urban economics.

By looking at the role of borrowing constraints in explaining neighborhood sorting, this paper belongs to a stream of literature that has considered several determinants of household sorting: differential preferences of races for single-family houses and/or highly-educated neighbors (Bajari & Kahn 2012); preferences for same-race neighbors (Durlauf 2004); education, language, immigration status (Bayer, McMillan & Rueben 2004); real estate brokers' steering behavior (Ondrich, Ross & Yinger 2003). The contribution of the paper is to introduce lending standards among the determinants of racial segregation in an equilibrium model.

The paper proceeds as follows. Section 2 introduces the economic model of location choice with borrowing constraints and its equilibrium properties. Section 3 introduces the empirical approach for identifying (section 3.1) and estimating (section 3.2) the model by simulated generalized method of moments. Empirical findings are described in Section 4. Section 5 presents the general equilibrium analysis of a change in lending standards. Section 6 discusses the impact of local housing supply elasticity and of tenure choice on estimates. Section 7 concludes the paper's findings.

2 The Model

We model a metropolitan area as composed by a number of neighborhoods, each formed by housing units, and inhabited by a corresponding population of households. Time is discrete and there is a finite number of time periods $t = 1, 2, \ldots, T$. In year t, the metropolitan area has a number N_t of households indexed by $i = 1, 2, \ldots, N_t$. There is a fixed number of $J \geq 2$ neighborhoods and each neighborhood $j = 1, 2, \ldots, J$ in year $t = 1, 2, \ldots, T$ has a number s_{jt} of housing units. Household i in year t chooses a neighborhood $j(i,t) \in \{1, 2, \ldots, J\} \equiv \mathbb{J}$, s.t. $j(i,t)$ belongs to the *choice set* $C_{it} \subset \mathbb{J}$ of household i in year t. The choice set C_{it} is a subset of $\mathbb{J}$ and is equal to $\mathbb{J}$ if and only if household i is unconstrained in year t.

(i) Neighborhood Choice $j(i,t)$ Conditional on Credit Availability

Household i derives utility U_{ijt} from neighborhood j in year t. Household i's neighborhood choice $j(i,t)$ is based on maximizing utility within the choice set C_{it}, i.e. $j(i,t) = \text{argmax}_{j \in C_{it}} U_{ijt}$. Utility U_{ijt} is allowed to vary across neighborhoods, across individuals, and to vary across neighborhoods according to individual characteristics. Thus:

$$U_{ijt} = \delta_{jt} + \boldsymbol{x}'_{it}\Omega\boldsymbol{z}_{jt} + \tilde{\boldsymbol{\beta}}'_{it}\boldsymbol{z}_{jt} + \varepsilon_{ijt} \tag{1}$$

All throughout the paper, bold symbols denote vectors. Household i's observable characteristics are stacked in a vector $\mathbf{x}_{it}$ of size $K_1 \geq 1$. Similarly, neighborhood j's observable characteristics are summarized by a vector $\mathbf{z}_{jt}$ of size $K_2 \geq 1$.[7] In utility specification (1), $\boldsymbol{\delta}_{jt}$ is the base utility that all households derive from living in neighborhood j in year t. The vector $\boldsymbol{\delta}_t$ thus captures variation in utility across neighborhoods independently of individual characteristics. $\mathbf{x}'_{it}\Omega\mathbf{z}_{jt}$ is a series of interaction terms between the K_1 household characteristics $\mathbf{x}_{it}$ and the K_2 neighborhood characteristics $\mathbf{z}_{jt}$, where Ω is the matrix of K_1K_2 interaction coefficients. Therefore $\mathbf{x}'_{it}\Omega\mathbf{z}_{jt}$ captures how neighborhood utility varies according to households' observable characteristics. In contrast, the term $\tilde{\boldsymbol{\beta}}'_{it}\boldsymbol{z}_{jt}$ captures how neighborhood utility varies in household-specific unobservable dimensions. $\tilde{\boldsymbol{\beta}}_{it}$ is a normally-distributed vector of household-level heterogeneity with mean 0 and with variance-covariance matrix Σ. Diagonal elements of Σ are noted σ_k, $k = 1, 2, \ldots, K_2$. Normalizing each random coefficient $\tilde{\beta}_{itk}$ by σ_k, with $\tilde{\boldsymbol{\beta}}_{it} = \left(\tilde{\beta}_{itk}\right)_{k=1,2,\ldots,K_2}$, the random term $b_{itk} = \tilde{\beta}_{itk}/\sigma_k$ is standard normal.

Base utility δ_{jt} is explained both by the neighborhood's log price, observable local amenities $z'_{jt}\beta$, and by unobservable neighborhood heterogeneity ξ_j:

$$\delta_{jt} = -\alpha \log(p_{jt}) + \mathbf{z}'_{jt}\beta + \xi_j + \zeta_{jt} \tag{2}$$

In utility specification (1), the random term ε_{ijt} is extreme-value distributed. Following McFadden (1973), household i's probability of choosing neighborhood j in year t, within his choice set C_{it}, is:

$$P\left(U_{ijt} \geq U_{ikt}, \forall k \in C_{it} \middle| \boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, \tilde{\boldsymbol{\beta}}_{it}, C_{it}\right) = \mathbf{1}(j \in C_{it}) \cdot \frac{\exp(\delta_{jt} + \boldsymbol{x}'_{it}\Omega\mathbf{z}_{jt} + \tilde{\boldsymbol{\beta}}'_{it}\mathbf{z}_{jt})}{\sum_{k \in C_{it}} \exp(\delta_{kt} + \boldsymbol{x}'_{it}\Omega\mathbf{z}_{kt} + \tilde{\boldsymbol{\beta}}'_{it}\mathbf{z}_{kt})} \tag{3}$$

where $\mathbf{1}(j \in C_{it}) = 1$ whenever $j \in C_{it}$, and 0 otherwise. Integrating over the vector of random

[7] $\mathbf{z}_{jt}$ can also include the characteristics of loans for location j for household i in year t. This is discussed in Section 4.3.

coefficients $\tilde{\boldsymbol{\beta}}_{it}$ obtains the probability conditional on observables and on base utility:

$$P(j,t|\boldsymbol{\delta}_{\cdot t},\mathbf{z}_{\cdot t},\mathbf{x}_{it},C_{it}) = \int_{\tilde{\boldsymbol{\beta}}} P(j|\boldsymbol{\delta}_{\cdot t},\mathbf{z}_{\cdot t},\boldsymbol{x}_{it},\tilde{\boldsymbol{\beta}},C_{it}) \cdot f(\Sigma^{-1/2}\tilde{\beta}) \cdot d\tilde{\beta} \tag{4}$$

where f is the density function of the standard i.i.d. multivariate normal of dimension K_2.

(ii) Endogenizing the Choice Set $C_{it} \subset \mathbb{J}$

This paper focuses on constraints on the choice set driven by mortgage credit approval decisions. A neighborhood j belongs to the choice set C_{it} of individual i in year t if a mortgage lender approves household i's application in neighborhood j.[8] We thus model the probability that a given neighborhood j belongs to the choice set C_{it} of household i by incorporating mortgage lenders' approval probabilities.

The choice set C_{it} is the product of banks' approval outcomes for the applications of household i. ϕ_{ijt} is the probability that household i is approved for a mortgage in neighborhood j in year t. With independent draws, the probability that household i's choice set C_{it} is exactly $C \subset \mathbb{J}$ is the binomial probability:

$$P(C_{it} = C) = \Pi_{j\in C}P(j \in C)\Pi_{j\notin C}\left[1 - P(j \in C)\right] \equiv \Pi_{j\in C}\phi_{ijt}\Pi_{j\notin C}(1-\phi_{ijt}) \tag{5}$$

Such formula assumes independent and identically distributed decisions, and an extension of the model allows for non-zero correlation in banks' decisions across neighborhoods for a given household i.[9]

The probability ϕ_{ijt} of approval results from banks' benefit-cost analysis of originating a mortgage loan for household i for a house in neighborhood j. Such benefit-cost analysis is summarized in a latent variable ϕ^*_{ijt}. Household i's application in neighborhood j is approved if and only if $\phi^*_{ijt} \geq 0$. The benefit-cost ϕ^*_{ijt} depends on household observables $\mathbf{x}_{it}$ and neighborhood characteristics $\mathbf{z}_{jt}$.

$$\phi^*_{ijt} = \mathbf{x}'_{it}\boldsymbol{\psi} + \mathbf{z}'_{jt}\boldsymbol{\gamma} + \mathbf{x}'_{it}\Psi\mathbf{z}_{jt} + e_{ijt} \tag{6}$$

[8] The endogeneity of the choice of mortgage lender for applications is addressed in Section 4.2.

[9] Results available from the authors. The updated formula 5 features the correlation of binomial draws across neighborhoods. Assuming independent approval probabilities will likely give us a lower bound of the impact of borrowing constraints.

Neighborhood characteristics $\mathbf{z}_{jt}$ include a neighborhood fixed effect that captures location-specific underwriting unobservables. The neighborhood's log price either enters directly in $\mathbf{z}_{jt}$, or enters as the log of the loan-to-income ratio in both $\mathbf{x}_{it}$ (log household income) and $\mathbf{z}_{jt}$ (log house price). In both cases, a change in log price affects the probability of the choice set. Ψ is the matrix of interaction terms. e_{ijt} is an i.i.d. logistically distributed random heterogeneity with cumulative distribution function $\Lambda(\cdot)$.[10] The probability of approval in neighborhood j for household i is therefore simply a logistic function of individual and household observables $P(j \in C_{it}) = \Lambda\left(\mathbf{x}'_{it}\boldsymbol{\psi} + \mathbf{z}'_{jt}\boldsymbol{\gamma} + \mathbf{x}'_{it}\Psi\mathbf{z}_{jt}\right)$. Combining (5) and such logit specification, the probability of the choice set depends on the vector of all neighborhood and household observables:

$$P(C_{it} = C|\mathbf{x}_{it}, \mathbf{z}_{\cdot t}) = \Pi_{j\in C}\Lambda\left(\mathbf{x}'_{it}\boldsymbol{\psi} + \mathbf{z}'_{jt}\boldsymbol{\gamma} + \mathbf{x}'_{it}\Psi\mathbf{z}_{jt}\right)\Pi_{j\notin C}\left\{1 - \Lambda\left(\mathbf{x}'_{it}\boldsymbol{\psi} + \mathbf{z}'_{jt}\boldsymbol{\gamma} + \mathbf{x}'_{it}\Psi\mathbf{z}_{jt}\right)\right\} \quad (7)$$

(iii) Neighborhood Demand

Total demand for neighborhood j in year t is the weighted average of the demands for neighborhood j conditional on all possible choice sets. The weights are equal to the corresponding choice set probabilities.

$$P(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}) = \sum_{C\in\mathbf{C}} P(j|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, x_{it}, C_{it}) \cdot P(C|\mathbf{z}_{\cdot t}, x_{it}) \quad (8)$$

As the set of all possible choice sets $\mathbf{C}$ has 2^J, such actual demand is the sum over 2^J terms. Equation (8) is the actual demand for the demographic subgroup of households with characteristics $\mathbf{x}_{it}$, i.e. the demand from households with a given income, race, education, age, and other characteristics. The last step in obtaining total demand, from all demographic subgroups, is to integrate over the distribution of individual demographics:

$$D(j,t) \equiv P\left(j,t\,|\boldsymbol{\delta}, \mathbf{z}_{\cdot t}, f(\cdot)\right) = \sum_{C\in\mathbf{C}} \int_{\mathbf{x}} P(j|\boldsymbol{\delta}, \mathbf{z}_{\cdot t}, x_{it}, C_{it}) \cdot P(C|\mathbf{z}_{\cdot t}, x_{it}) \cdot f(\mathbf{x}_{it}) d\mathbf{x}_{it} \quad (9)$$

And $f(\mathbf{x}_{it})$ is the city-level distribution of household characteristics. We note $D(j,t)$ such total demand for neighborhood j in year t.

[10] The paper's section 4.2 also considers a normal distribution for e_{ijt} when instrumenting the covariates.

(iv) Demand Elasticities

As each household demand depends on the particular choice set probabilities of the household, demand elasticities conditional on the choice set differ from demand elasticities unconditional on the choice set. The first kind of (own-price) demand elasticity $\eta_{|S}$ is equal to the effect of a change in the log price on demand while keeping the choice set probabilities $P(C_{it})$, $i = 1, 2, \ldots, N$, $t = 1, 2, \ldots, T$ constant. For a fixed probability distribution $P(C)$ over choice sets, we note $\eta_{jt|C}$ the own-price demand elasticity for neighborhood j in year t conditional on $P(C)$.

$$\begin{aligned}
\eta_{jt|C} &= \frac{1}{D(j)} \left.\frac{\partial D(j,t)}{\partial \log(p_{jt})}\right|_{P(C)\,fixed} \qquad (10)\\
&= -\frac{1}{D(j,t)} \sum_{C \in \mathbf{C}} \int_X \alpha(\mathbf{x}_{it}, \tilde{\boldsymbol{\beta}}_{it}) \cdot P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}) P(j,t|\boldsymbol{\delta}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C)\left(1 - P(j,t|\boldsymbol{\delta}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C)\right) f(\mathbf{x}_{it}) d\mathbf{x}_{it}
\end{aligned}$$

where $\alpha(\mathbf{x}_{it}, \tilde{\boldsymbol{\beta}}_{it})$ is the impact of log price on base utility for a household with characteristics $\mathbf{x}_{it}$ and random coefficient $\tilde{\boldsymbol{\beta}}_{it}$ (Equation (2)). In the simplified case where the model does not feature random coefficients, and when there is a large number of neighborhoods so that $P(j) \ll 1$, such conditional demand elasticity is simply $-\alpha$.

When prices change, probabilities $P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it})$ over choice sets $C \subset \mathbb{J}$ change as well. Total demand elasticity η is equal to the sum of two effects: the price effect on demand due to utility changes, and the price effect on demand due to changes in choice set probabilities,

$$\eta_{jt} = \eta_{jt|C} + \sum_{C \in \mathbf{C}} \int_{\mathbf{x}} \left[\frac{\partial}{\partial \log(p_{jt})} P(C|\mathbf{z}_{\cdot t}, x_{it}) \right] \cdot P(j,t|\boldsymbol{\delta}, \mathbf{z}_{\cdot t}, x_{it}, C_{it}) \cdot f(\mathbf{x}_{it}) d\mathbf{x}_{it} \qquad (11)$$

The second term of this expression (11) is the second price effect. An increase in house price in neighborhood j causes a decline in demand because of (i) the lower utility value of the neighborhood and (ii) the lower approval rate in neighborhood j.

The impact of a *log* price change on the choice-set probability is simply derived from the individual approval probabilities.

$$\begin{aligned}
\frac{\partial}{\partial \log(p_j)} P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi) &= \frac{\partial}{\partial \log(p_j)} \left\{ \Pi_{k \in C} \phi_{ijt} \Pi_{k \notin C} (1 - \phi_{ijt}) \right\} \\
&= a \cdot \underbrace{(\mathbf{1}(j \in C) - \phi_{ijt})}_{\Phi_{ijt}(C)} P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}) \qquad (12)
\end{aligned}$$

where a is the coefficient of *log* price in the approval specification (6). The closed-form expression of the borrowing elasticity follows by plugging (12) into (11).

(v) City Equilibrium

Households choose their location based on neighborhoods' observable characteristics, including the log price and neighborhoods' characteristics. Therefore, two sets of conditions need to be satisfied at equilibrium.

The first set of conditions is the equality of demand and supply for each neighborhood at equilibrium. Subsections (i)–(iv) above have introduced household demand and its properties. The supply of housing in neighborhood j depends on the log price of housing of housing in neighborhood j such that the supply of housing $S_j(p_{jt})$ in neighborhood j is an increasing and concave function of p_{jt}. In this paper's empirical analysis, Sections 4–5, supply is initially assumed perfectly inelastic, so that $S_j(p_{jt}) = s_{jt}$ for every j, t.

Households have preferences for specific neighborhood demographics whenever the vector $\mathbf{z}_{jt}$ of neighborhood characteristics includes other households' demand for the same neighborhood. Neighborhood characteristics thus comprise a set $\mathbf{w}_{jt}$ of exogenous characteristics (e.g. the age of structures in the neighborhood) and a set of neighbors' demographics $\mathbf{v}_{jt}$ (e.g. fraction of households with a college degree in neighborhood j). Note $v_{jt}(x)$ the fraction of households with characteristic x. Write X^+ the set of neighbors' characteristics that affect household utility, then $\mathbf{v}_{jt} = (v_{jt}(x),\ x \in X^+)$.

Definition 1. *Given the distribution of household observables $f(x_{it})$, the vector of neighborhood observables $\mathbf{z}_{jt}$, excluding log price, and unobservables $\boldsymbol{\xi}_j$, $\boldsymbol{\zeta}_{jt}$, $j = 1, 2, \ldots, J$, $t = 1, 2, \ldots, T$, a* city equilibrium *is a vector of prices and neighborhood demographics $(\mathbf{p}^* = (p^*_{jt}), \mathbf{v}^* = (\mathbf{v}^*_{jt}))_{j=1,2,\ldots,J,\, t=1,2,\ldots,T}$ such that:*

$$\forall j \in \mathbb{J}, \forall t \in \{1, 2, \ldots, T\}, \quad D\left(j, t; \mathbf{p}^*, \mathbf{v}^*\right) = S_j\left(p^*_{jt}\right) \tag{13}$$

$$\forall x \in X^+, \quad \mathbf{v}_{jt}(x) = \frac{D(j, t|x; \mathbf{p}^*, \mathbf{v}^*)}{D(j, t; \mathbf{p}^*, \mathbf{v}^*)} \tag{14}$$

Condition (13) expresses the equality of demand and supply. In condition (14), $\mathbf{v}_{jt}(x)$ is the fraction of households with characteristic x, for all neighborhood demographics x that enters households' *utility for neighborhood j; $D(j,t|x;\mathbf{p}^*,\mathbf{v}^*)$ is household xs' demand for neighborhood j in year t. $\frac{D(j,t|x;\mathbf{p}^*,\mathbf{v}^*)}{D(j,t;\mathbf{p}^*,\mathbf{v}^*)}$ is the fraction of such households as derived from equilibrium neighborhood demand.*

The Appendix's proposition 1 proves equilibrium existence using Brouwer's theorem. Proposition 2 proves global equilibrium uniqueness when households do not exhibit preferences for peers' demographics, using an argument of gross substitutability. An equilibrium is locally unique as in Debreu (1970), proven using Sard's theorem. Local uniqueness enables comparative statics analysis of lending standards (Section 5).

3 Identification and Estimation

This section presents the empirical approach for the identification (section 3.1) and estimation (section 3.2) of borrowing constraints and of households' preferences for neighborhood amenities and neighborhood demographics.

3.1 Identification Strategy

The model includes three sets of parameters to estimate, which we identify using three sets of corresponding moments. The first set of parameters determines the probability of approval for a mortgage application, as given by the approval logit specification. This paper identifies such parameters using an instrumental variable strategy. The second set of parameters measures the impact of neighborhood observable and unobservable characteristics on base utility, as in specification (2). Identifying the impact of log house price on base utility requires a set of instruments: log house prices are typically correlated with time-varying unobservables that are not controlled for by the neighborhood-specific fixed effect. The paper presents a set of IV moments as in by Berry et al. (1995) to that effect. Finally, the third set of parameters measures how the valuation of neighborhood amenities depends on household characteristics. We identify such preference heterogeneity using micromoments as in Imbens & Lancaster (1994) and Petrin (2002).

(i) First Set of Moments: Lending Standards

The approval specification (6) predicts the probability of approval based on the characteristics of the household, the time-varying characteristics of the neighborhood, a neighborhood fixed effect, and interactions between individual and neighborhood characteristics. The neighborhood fixed effect captures non time-varying unobservables that are correlated with the unobservable collateral (house) value for the loan and have an impact on the probability of approval.

The identification challenge lies (i) in the lack of observability of the full range of time-varying household and house covariates that mortgage originators consider in the underwriting process and (ii) in the endogeneity of the choice of mortgage lender. Empirical Section (4.2) presents a set of instrumental variables $\mathbf{\Theta}_{ijt}$ that are arguably uncorrelated with the unobservables $\mathbf{e}_{ijt}$ of specification (6), affect the probability of approval, and predict the characteristics $\mathbf{x}_{it}$ of mortgage applicants. Such instrumental variables, used in the context of a logit, provide the first set of moment conditions:

$$E\left[G_0\left(\psi_0, \gamma_0, \omega_0\right)\right] = E\left[\Lambda\left(\mathbf{x}_{it}'\psi + \mathbf{z}_{jt}'\gamma + \mathbf{x}_{it}'\Psi\mathbf{z}_{jt}\right)\mathbf{\Theta}_{ijt}\right] = 0 \tag{15}$$

(ii) Second Set of Moments: Base Utility

The second set of moments aims at identifying the impact of neighborhood amenities and neighborhood demographics on neighborhood base utility (Specification (2)). Identification of such parameters $(\alpha, \boldsymbol{\beta}, \boldsymbol{\xi})$ encounters at least three challenges (i) neighborhood utility U_{ijt} in specification (1) is not directly observable, and thus base utility δ_{jt} at the left-hand side of specification (2) is not directly observable; (ii) demand, defined in equation (9), is the sum of a large number of 2^J terms, one for each potential choice set $C_{it} \subset \{1, 2, \ldots, J\}$, (iii) unobservable neighborhood amenities ζ_{jt} may be correlated with the price of housing and with observable neighborhood amenities (e.g. school quality).

On the first issue, Berry et al. (1995) has shown that, given the interaction parameters Ω, and the variance of the random coefficients Σ, there is a one-to-one mapping between the vector of observed demands $\mathbf{D} = (D_{jt})_{j\in\mathbb{J}, t\in\{1,2,\ldots,T\}}$ and the vector of base utilities $\boldsymbol{\delta} = (\delta_{jt})_{j\in\mathbb{J}, t\in\{1,2,\ldots,T\}}$, whenever households are unconstrained, i.e. $\phi_{ijt} = 1$ for all households, all neighborhoods and all years. A similar result of existence and uniqueness of the vector $\boldsymbol{\delta}$ applies when demand is the convex weighted average of the demand given each choice set, with probabilities $\phi_{ijt} \in [0, 1]$ of the choice sets. The unique vector $\boldsymbol{\delta}$ is obtained by iterating the following sequence:

$$\hat{\boldsymbol{\delta}}^{k} = \hat{\boldsymbol{\delta}}^{k-1} + \log\left(D_{jt}\right) - \log\left(D\left(j,t|\hat{\boldsymbol{\delta}}^{k-1}\right)\right) \tag{16}$$

and at the limit $\hat{\boldsymbol{\delta}} = \lim_{k\to\infty} \hat{\boldsymbol{\delta}}^{k}$ of such sequence, predicted demand $D\left(j,t|\lim_{k\to\infty}\hat{\boldsymbol{\delta}}^{k}\right)$ is equal to observed demand D_{jt}.

In the contraction mapping (16) , demand $D\left(j,t|\hat{\boldsymbol{\delta}}\right)$ is the sum of demand for all possible choice sets $C \subset \{1,2,\ldots,J\}$, weighted by the probability of that choice set $P(C)$. With J neighborhoods, there are 2^{J} choice sets. In our empirical application where households choose among more than 4000 neighborhoods, computing demand over a sum of 2^{4000} choice sets is unfeasible. We proceed by simulating $D\left(j,t|\hat{\boldsymbol{\delta}}\right)$. A set of $\mathcal{S}$ sets is drawn C_{it}^{s}, $s = 1,2,\ldots,\mathcal{S}$ for each household i in each year t, where the probability of drawing the set $C \subset \mathbb{J}$ of neighborhoods is $\Pi_{j\in C}\hat{\phi}_{ijt}\Pi_{j\notin C}\left(1-\hat{\phi}_{ijt}\right)$. Estimation of demand $D\left(j,t|\hat{\boldsymbol{\delta}}\right)$ given base utilities, and observables thus is the average of the demands based on the simulated choice sets C_{it}^{s}:

$$\hat{D}\left(j,t|\hat{\boldsymbol{\delta}}\right) = \sum_{i=1}^{N}\frac{1}{\mathcal{S}}\sum_{s=1}^{\mathcal{S}} D\left(j,t|\hat{\boldsymbol{\delta}},C_{it}^{s}\right) \tag{17}$$

with $D\left(j,t|\hat{\boldsymbol{\delta}},C_{it}^{s}\right)$ the demand conditional on the choice set C_{it}^{s}. Train (2009) gives conditions under which such a simulated method yields a consistent estimate of demand. Furthermore, the integration over household characteristics $\int_{\mathbf{x}} f(\cdot)dx$ and over the random coefficients $\int_{\tilde{\boldsymbol{\beta}}} f(\cdot)d\tilde{\boldsymbol{\beta}}$ cannot proceed using analytic formulas and so we rely on Monte Carlo integration following Geweke (1996). The simulated demand $\hat{D}$ thus replaces the actual demand D in the sequence ($\boldsymbol{\delta}^{k}$) specified in (16).

The contraction mapping method provides an estimate $\hat{\delta}_{jt}$ for each neighborhood in each year. The estimation of the regression (2) by least squares will not identify the impact of prices and amenities on utility. Least squares regression typically yield a positive coefficient on log price, while consumer theory suggests a negative coefficient $-\alpha$. The main reason is that higher prices are likely to be correlated with the unobservable quality of the housing stock, possibly leading to an upward bias on the log(price) coefficient in households' utility function. Empirical section 4.3 introduces L' time-varying instruments $\boldsymbol{\Xi} = \Xi_{jt}$ orthogonal to the unobservable neighborhood amenities ζ_{jt}, providing a set of moment conditions for the estimation of the impact of observable neighborhood amenities, house value, neighbors' demographics $\mathbf{z}_{jt}$ on base utility δ_{jt}. Specification (2) relating base-utility to house and neighborhood characteristics $\mathbf{z}_{jt}$ is first-differenced. The first moment condition expresses the orthogonality of $\Delta\boldsymbol{\Xi}$ with $\Delta\boldsymbol{\zeta}$:

$$E(\Delta\boldsymbol{\Xi}\Delta\boldsymbol{\zeta}) = 0 \tag{18}$$

where the change in unobservable amenities $\Delta\zeta_{jt}$ depends on observable neighborhood characteristics: $\Delta\zeta_{jt} = \Delta\delta_{jt} - \Delta z_{jt}^{'}\beta$, whose estimator $\Delta\hat{\zeta}_{jt} = \Delta\hat{\delta}_{jt} - \Delta z_{jt}^{\prime}\beta$ follows from the previously estimated vector of base utilities $\hat{\boldsymbol{\delta}}$.

$\Delta\boldsymbol{\zeta}$ depends on lending standards $(\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi)$ and on preference parameters $(\alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma)$ as the contraction mapping depends on predicted approval probabilities $\hat{\phi}_{ijt}$ and on the non linear parameters (Ω, Σ). We write the moment condition (18) as a function $E\left[G_1(\theta)\right] = E\left[\Delta\boldsymbol{\Xi}\Delta\boldsymbol{\zeta}(\boldsymbol{\theta})\right] = 0$ of the vector $\boldsymbol{\theta}$ of lending standards and preference parameters.

Moment condition (18) together with demand simulation (17) and the contraction mapping (16) allows for the estimation of households' preferences for neighborhood amenities, i.e. the coefficient β, but households' preference heterogeneity remains to be estimated and is covered in the next subsection.

(iii) Third Set of Moments: Preference Heterogeneity

The matrix of interaction coefficients Ω in utility specification (1), measures how heterogeneity in household preferences is explained by households' observables, and the variance covariance matrix $\Sigma = E\left[\tilde{\boldsymbol{\beta}}_{it}\tilde{\boldsymbol{\beta}}_{it}^{\prime}\right]$ measures how unobservable household dimensions explain preference heterogeneity. Identification of the parameters Ω and Σ in (1) is provided by a third set of micro moment conditions following Imbens & Lancaster (1994) and Petrin (2002). Specifically, the third set of micro moments ensures that the predicted spatial distribution of households, e.g. by race and income, matches the observed distributions of such households across neighborhoods.

Two data sources provide moment conditions useful for such identification. First, Census micro data provides the characteristics $\mathbf{x}_{it}$ of a sample of size N for all households $i = 1, 2, \ldots, N$. Second, neighborhood-level data provides elements of the distribution of household characteristics x in each neighborhood $j \in \mathbb{J}$. For instance, the Census provides log median family income per neighborhood (blockgroup in the empirical section). Such information provides a third set of moments as we match the predicted demand by demographic subgroup for each neighborhood to the actual demand by this same subgroup. Consider $\mathcal{X}$ a set of observable characteristics such as, for example, the set of households with income above \$50,000. Then the following moment states that the number of households with characteristics $\mathcal{X}$ predicted by the model must be equal to the number of such households in each neighborhood.

$$E\left(\sum_{x_{it}\in\mathcal{X}} D\left(j|\hat{\boldsymbol{\delta}}_{\cdot t}, \mathbf{z}_{\cdot t}, x_{it}, C_{it}\right)\right) - D_{jt}^{\mathcal{X}} = 0 \tag{19}$$

with $D_{jt}^{\mathcal{X}}$ the observed number of such households in neighborhood j in year t.

The sum $\sum_{x_{it}\in\mathcal{X}} D\left(j, t|\hat{\boldsymbol{\delta}}_{\cdot t}, \mathbf{z}_{\cdot t}, x_{it}, C_{it}\right)$ is the demand for neighborhood j from individuals with characteristics in $\mathcal{X}$ as predicted by the model. A number L' of subsets $\mathcal{X}_1, \ldots, \mathcal{X}_{L'}$ of households provides a set of L' moment conditions, $E\left[G_2\left(\boldsymbol{\theta}\right)\right] = 0$.

(iv) Estimation using the Three Sets of Moments

Lending standards $(\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi)$ are estimated jointly with preferences $(\alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma)$. We stack the L instrumental variable moment conditions of the approval specification, the L' moment conditions of base utility analysis, and the L'' moment conditions for subpopulations. Writing G as (G_0, G_1, G_2), parameters satisfy:

$$E\left(G(\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi, \alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma)\right) = 0 \tag{20}$$

A consistent and asymptotically normal estimator of household preferences and banks' lending standards follows from Hansen (1982) as the minimand $\hat{\boldsymbol{\theta}} = \text{argmin}_{\theta\in\Theta} G^*(\boldsymbol{\theta})' G^*(\boldsymbol{\theta})$, where $\boldsymbol{\theta} = (\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi, \alpha, \boldsymbol{\beta}, \boldsymbol{\xi}, \Omega, \Sigma)$ and $G^*(\boldsymbol{\theta}) = A(\tilde{\boldsymbol{\theta}})\hat{G}(\boldsymbol{\theta})$, and A is a consistent estimate of the square root of the inverse of the asymptotic variance-covariance matrix of the moments, obtained using $\tilde{\boldsymbol{\theta}}$, a preliminary consistent estimate of $\boldsymbol{\theta}$. $\hat{G}(\boldsymbol{\theta})$ is the simulated sample analogue of $G(\boldsymbol{\theta})$.

3.2 Estimation Steps

The model is estimated using three sources of data: the mortgage application and approval data, the neighborhood data, and the household micro data. As the first set of moments is taken from a different sampling process than the two other moments, estimating the lending standards $\hat{\boldsymbol{\psi}}$, $\hat{\boldsymbol{\gamma}}$,$\hat{\Psi}$ can be performed separately in a first step. We then proceed in the following steps for the estimation of household preferences $\hat{\alpha}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\xi}}, \hat{\Omega}, \hat{\Sigma}$:

1. For each observation i, t of the household sample:

   (a) Predict the probability of approval $\hat{\phi}_{ijt}$ for each neighborhood j using the estimate $\Lambda\left(\mathbf{x}_{it}'\hat{\boldsymbol{\psi}} + \mathbf{z}_{jt}'\hat{\boldsymbol{\gamma}} + \mathbf{x}_{it}'\hat{\Psi}\mathbf{z}_{jt}\right)$.

   (b) Draw $\mathcal{S}$ choice sets for individual i by drawing a dummy variable for each neighborhood j equal to 1 with probability $\hat{\phi}_{ijt}$. This defines a matrix of 0s and 1s, noted C_{ijt}.

   (c) Draw a vector of i.i.d. multivariate standard normal random coefficients $\tilde{\mathbf{b}}_{it}$.[11]

2. Then we minimize the GMM objective function $G^*(\boldsymbol{\theta})'G^*(\boldsymbol{\theta})$. To obtain $G(\boldsymbol{\theta})$ for a given vector of parameters $\boldsymbol{\theta}$:

   (a) Estimate the vector $\hat{\boldsymbol{\delta}}_t$ of base utilities using the contraction mapping (16).

   (b) At each iteration of the contraction mapping, simulated total demand is $\hat{D}(j,t) = \frac{1}{N_t}\sum_{i=1}^{N_t}\hat{D}(j,t|\mathbf{x}_{it}, \mathbf{z}_{jt}, \hat{\boldsymbol{\delta}}_t, \tilde{\boldsymbol{\beta}}_{it})$, the average of individual demands. And simulated individual demand is $\hat{D}(j,t|\mathbf{x}_{it}, \mathbf{z}_{jt}, \hat{\boldsymbol{\delta}}_t, \tilde{\boldsymbol{\beta}}_{it}) = \mathbf{1}(C_{ijt}=1)\cdot\frac{\exp(\delta_{jt}+\boldsymbol{x}_{it}'\Omega\mathbf{z}_{jt}+\tilde{\boldsymbol{\beta}}_{it}'\mathbf{z}_{jt})}{\sum_{k,s.t.C_{ikt}=1}\exp(\delta_{kt}+\boldsymbol{x}_{it}'\Omega\mathbf{z}_{kt}+\tilde{\boldsymbol{\beta}}_{it}'\mathbf{z}_{kt})}$. The $\tilde{\boldsymbol{\beta}}_{it}$ are obtained by multiplying $\tilde{\mathbf{b}}_{it}$ by $\Sigma^{1/2}$.

   (c) Perform the panel instrumental variable regression of base utilities $\hat{\boldsymbol{\delta}} = (\hat{\boldsymbol{\delta}}_t; t = 1, 2, \ldots, T)$ on neighborhood covariates instrumented by the vector Ξ, to obtain the panel residuals $\Delta\hat{\zeta}(\boldsymbol{\theta})$ and compute the second empirical moments $\Delta\Xi\Delta\hat{\zeta}(\boldsymbol{\theta})$.

   (d) For each demographic subgroup $\mathcal{X}$, estimate the demand of that subgroup for each neighborhood using step (b)'s individual demands. This gives the second set of moment conditions.

   (e) Steps (c) and (d) together give the last two moments G_1 and G_2 of the objective function $G(\boldsymbol{\theta})$. The weighting matrix is diagonal in the first step estimation, and, in the second step, equal to an estimate of the inverse of the variance covariance matrix of the moments based on the first step estimate $\hat{\boldsymbol{\theta}}$.

[11] We use antithetic acceleration as in Geweke (1988) and Goeree (2008).

Further details of the optimization algorithm are presented in Appendix section B.

4 Empirical Findings

4.1 Data

The model is estimated for the 9 counties which are contiguous to the San Francisco Bay: Alameda, Contra Costa, Marin, Napa, San Francisco, San Mateo, Santa Clara, Solano, and Sonoma. Data is gathered for three decades: in 1990, 2000, and 2010.[12]

Mortgage Application Data

Mortgage application and approval data comes from data collected in accordance with the Home Mortgage Disclosure Act (HMDA). The act mandates reporting of mortgage application data by most depository and non-depository lending institutions.[13] Thus, HMDA disclosure requirements apply to more than 90% of all mortgage applications and originations (Dell'Arriccia, Igan & Laeven 2009). We focus on credit standards for first lien mortgages on homeowner occupied houses.[14] Each mortgage lender reports the loan amount, the applicant's income, the applicant's race and gender, and the census tract of the house. Such geographical information allows this paper to present lending standards that are location-specific and that vary across the distribution of household income and across races.

HMDA data are geographically matched to data on individual property transactions. Data on individual property transactions was obtained from the mortgage company FNC Inc.,[15] which compiles transaction data based using public records and real estate sales. FNC data reports the transaction price and the street address for the complete universe of transactions. Street addresses are geocoded using the transaction's street address and linked to each of our census tracts.[16]

[12] As the borders of geographies such as blockgroups and tracts change over time, we use constant 2000 borders throughout the analysis, i.e. for the three decennial waves of 1990 to 2010.

[13] Specifically, HUD regulates for-profit lenders that have combined assets exceeding $10 million and/or originated 100 or more home purchase loans (including refinancing loans) in the preceding calendar year.

[14] HMDA data contains information on the seniority of mortgage liens only starting in 2004. To eliminate second lien mortgages, also known as piggyback loans, throughout the sample, the dataset includes applications with a minimum LTI of 1.8; such threshold, according to the 2004 HMDA data, eliminates about 96 percent of second lien mortgages, while discarding only 5 percent of first lien mortgages.

[15] FNC Inc., headquartered in Oxford Mississippi, collects data that provides collateral information to the mortgage industry.

[16] Geocoding was performed using Texas A&M's geocoding services, with a higher than 90% success rate.

Housing and Amenities Data

Households choose dwelling location based on the quality of housing structures, local amenities, transportation, and neighborhood demographics.

Neighborhood-level data is derived from Census block-group data of the 1990 Summary Files 1 and 3, the 2000 Census Summary File 1, and the 2010 Summary File 1.[17] There are 4,418 block groups covering the San Francisco Bay area, with a median of 486 housing units per block group, and a median land area of 0.57 square mile. The block group is the smallest geographical unit for which the Census provides housing stock and demographic characteristics across three decades. As the borders of block groups may change across the three decennial census waves, we keep constant 2000 block group borders by building new block group-level census relationship files. Neighborhood characteristics are thus made comparable across the three waves.

The housing and amenities data comprises information on three dimensions that affect households' choice: characteristics of the housing structures (age of structure, median number of rooms, price), school test score data, and neighborhood demographics (race, ethnicity, household median income, fraction with college or associate degree, fraction with high school or more).

California's Department of Education provides Academic Performance Index (API) test score data for each school of the Bay area, for the last two waves, 2000 and 2010 of our neighborhood data. We use school district borders from the Census Bureau to match a block group to each of the 141 elementary or unified school districts, and assign the elementary school API to each block group. APIs are standardized by year to a mean of 0 and a standard deviation of 1.

For the distribution of house values, we choose to use FNC transaction price data rather than Census house values because, unlike the former, the latter are upper-censored and self-reported, with the upper-censoring threshold depending on the census year.[18] We verify that the distribution of FNC transaction prices matches well the distribution of Census house values up to the censoring points. With 126,104 transactions in 2000, a typical blockgroup of the San Francisco Bay Area has about 27.6 transactions in 2000. Thus, for each block group, we built a house value measure using

[17] We thank the National Historical Geographic Information System at the University of Minnesota for providing formatted Census files.

[18] Census values comprise the entire universe of houses, while FNC transaction data only measures prices for the subset of houses that transact. However, census values are self-declared and may thus be subject to non-classical measurement error.

the median transaction price[19] in each block group and for each year of observation.

Block group data also includes a measure of the distance to the San Francisco Central Business District as a proxy for access to retail and employment,[20] and land area in square miles to account for varying land surfaces across block groups.

Household data

This paper's third data source is the 5% Census sample in 1990 and 2000, and the American Community Survey in 2010, which are representative samples of households of the San Francisco Bay area.[21] Household-level data provides an empirical estimate of the distribution $f(\mathbf{x}_{it})$ of household characteristics. It thus allows for the integration of demand with respect to household characteristics $\int df(\mathbf{x})$, as preferences for housing, local amenities, and neighbors' characteristics depend on individual household characteristics (third moment of the model's identifying conditions (20)). Individual characteristics include race (White, Black, Hispanic, Asian),[22] and income as measure by total money income of all household members age 15 or above during the previous year.

A representative sample of households also allows us to predict approval probabilities for *all* household mortgages using estimated lending standards, regardless of whether they applied for a mortgage in a given year or not (first moment of the model's identifying conditions (20)). In order to do so, we use the fact that the variables used to predict approval probabilities for mortgage applications can be obtained for *all* households by combining representative household data from the Census and transaction data from FNC.

4.2 Estimation of Approval Probabilities

This section's goal is to estimate the approval regression model (6) using actual mortgage approval decisions. These estimates are then used to predict approval probabilities for each household of the

[19] We also estimated the 25th percentile and the 75th percentile of transaction prices. Although the heterogeneity of housing within a neighborhood could be a potential identification issue, estimation including such moments in base utility regression did not reveal that such heterogeneity plays a substantial role in determining utility.

[20] Borders of the CBD for the San Francisco bay area are according to the 1982 Census of Retail Trade as in Glaeser, Gottlieb & Tobio (2012).

[21] We acknowledge support from the Integrated Public Use Microdata Series center at the University of Missouri for providing comparable micro series across three decades.

[22] We use consistent definitions for races, from the 2000 Census and as defined by the Office of Management and Budget's 1997 Revisions to the Standards for the Classification of Federal Data on Race and Ethnicity. Thus a household is either "White, non-Hispanic," "Black, non-Hispanic," "Hispanic, of any race," "Asian, non-Hispanic," or of "Any other race, non-Hispanic."

micro household data, across all 4,418 neighborhoods. HMDA (described in the previous section) is, to our knowledge, the only data source that informs approval decisions for the universe of mortgage applications in the metropolitan area. HMDA data report, for each application, the approval decision, the income and the race of the applicant, the size of the loan,[23] the tract-level location of the application, and the financial institution where the application was filed. Other mortgage data sources, such as Dealogic or Black Box, report more information such as the loan-to-value ratio, the credit score, or the transaction price, but only for the universe of originated mortgages loans.

Identification Strategy

A regression of the mortgage originator's approval or denial decision on mortgage loan, household, and housing characteristics will typically not yield causal estimates of the impact of such characteristics on the mortgage originator's decision. At least two endogeneity issues bias such estimates based on observational data: first, the characteristics of the loan and of the household are correlated with unobserved determinants of the approval and denial decision, which include, but are not restricted to, the applicant's creditworthiness, and the house's unobservable quality dimensions. Second, households tend not to choose randomly the financial institutions to which they apply for a loan. They rather select banks endogenously so as to maximize the likelihood of obtaining the type of mortgage loan they need.[24] Therefore unobservable dimensions of lending standards are correlated with the observable loan, household, and housing characteristics.

Figure 1 shows, in 1994, the distribution of bank branches for a specific subarea of the San Francisco Bay Area.[25] The map displays only branches of banks regulated by the Federal Reserve System, the Office of the Comptroller of the Currency (OCC) and the Federal Deposit Insurance Corporation (FDIC), as these are mortgage originators for which balance sheet information is available from Reports of Income and Condition data. The location of bank branches is obtained in the Summary of Deposits data set provided by the Federal Deposit Insurance Corporation. As

[23] Post 2002 HMDA data also report the interest rate of the loan.

[24] Households could also file multiple applications for the same house. In practice however the observed number of applications per household is small. According to the estimates of Ouazad & Rancière (Forthcoming), the average U.S. household that changes house files between 2 and 3 applications.

[25] The map focuses on a specific geographic subarea. This paper's dataset covers all counties adjacent to the Bay area.

our mortgage application data set includes the census tract of the mortgage application's house, we match each tract to the five geographically closest bank branches.[26] But the location of bank branches may itself be endogenously driven by local demand, as bank branches may be opened or closed depending on local demand for mortgages. Therefore we also consider the case in which bank branches set up after 1985 are excluded. As a first step towards solving endogeneity issues, our regression thus uses the average loan-to-income ratio of applications of the five bank branches (excluding the considered loan-to-income ratio of the application if needed) as an instrument for the loan-to-income ratio of the mortgage application.

Of course, if applicants sort across neighborhoods endogenously, the average loan-to-income ratio of the five closest bank branches may address the endogeneity of bank selection but not the endogeneity of household characteristics. In the spirit of Loutskina & Strahan (2009), we use the *liquidity* level of the banks of the five closest bank branches as an instrument for the loan-to-income ratio. The idea underlying this approach is that liquidity, measured on the national balance sheet of a bank, is exogenous to its branch-level credit market conditions (or borrowers' characteristics), but nevertheless correlated with its general lending standards policy.

Figure 1 colors bank branches according to their banks' liquidity. Liquidity is derived following Loutskina & Strahan (2009) and Loutskina (2011) as the ratio of securities over total assets obtained from the Federal Reserve of Chicago's Reports of Income and Condition (the Call Reports). Red dots indicate branches whose national bank has low levels of liquidity while blue squares indicate branches whose national bank have high levels of liquidity. The map likely suggests that bank branches with high or low liquidity levels are not specifically located in high- or low-income neighborhoods, or white, African-American, Hispanic, or Asian neighborhoods.

Results

In the baseline logit specification,[27] we predict the decision of mortgage approval, for the panel of mortgage applications filed in 1990, 2000, and 2010, based on a number of borrower and loan

[26] Bank branches' latitude and longitude is included in the Summary of Deposits. We consider the centroid of each census tract from Census Bureau shapefiles. Then each tract centroid is matched to the five bank branches that are closest. The projected coordinate reference system is the North American Albers Equal Area Conic.

[27] Using linear instrumental variable regression (with or without fixed effect) yields coefficients that are similar to the marginal probability effects of the probit and logit specification. However linear methods' predictions typically fall outside the $[0, 1]$ interval with non zero probability and cannot be used to estimate choice probabilities.

characteristics, including the loan-to-income ratio (LTI), the race of the applicant, the type of loan (Conventional, FHA insured, VA insured, FSA-RHS insured[28]), time effects, and census-tract fixed effects. Results are presented in Table 1. The upper part of the table shows results without and with census-tract fixed effects in the logit specification. Appendix Table B shows similar results with a probit estimation. We use probit for the instrument variable estimation (bottom part of Table 1). Probit instrumental variable's moment conditions follow equation (15) and IV probit estimation is readily available in common statistical packages.[29] For each estimation, the marginal probabilities, computed at the means, and standard errors computed using a delta method, are presented next to the logit coefficients. The coefficients of the race dummies (resp., the type of loan) should be interpreted as relative to the dummy for white households (resp., relative to the dummy for Conventional loan). Standard errors are clustered at the census tract level.

In uninstrumented logit and probit specifications, the coefficients for the LTI Ratio and for race dummies are significant at one percent confidence level. As expected, the LTI ratio coefficient is negative: applying for a larger loan relative to the applicant's income reduces the chances of obtaining approval. A one-standard deviation increase in LTI (+1.28) is associated in a reduction in the probability of approval ranging from 2.7 to 3.7 percentage points, depending on the specification. Conditional on the observable characteristics included, black and hispanic applications face a lower probability of approval. This negative effect gets mitigated as the census tract fixed effect is introduced.

The bottom panel of Table 1 presents the results of the probit estimation when instrumenting the loan-to-income ratio of the application. The first two columns present results using the average loan-to-income ratio of the applications of the five geographically closest bank branches. The loan-to-income ratio of the application itself is excluded from the average. The next two columns (Columns 3 and 4) present our preferred estimates, i.e. results where the loan-to-income ratio is instrumented by the average liquidity of the corresponding national banks, for the five geographically-closest branches. The sample is made of observations for which we can obtain liquidity measures for banks and for which the geographic identifier (census tract) can be matched to corresponding census data.

[28]FHA - Federal Housing Administration; VA - Veteran Administration; FSA-RHS-(Farm Service Agency or Rural Housing Service

[29]The instrumental variable probit model is presented in Newey (1987).

[30] Appendix Table C presents robustness tests of the IV strategy. Considering only bank branches that were set up prior to 1985 (first two columns), thus alleviating the potential endogeneity of branching decisions, produces comparable results. Similar results are also obtained when, in addition to instrumenting the LTI, we also instrument the race of the applicant (indicator variables for black, Hispanic, asian applicant) by the racial composition of neighboring tracts (last two columns).

In all IV regressions (bottom panel of Table 1 and Appendix Table C), the F statistics for each first stage, for both the loan-to-income and race indicator variables, are well above 10 and Cragg Donald finite sample bias F statistics are substantially above the threshold for a maximum 5% finite sample bias; this suggests that finite sample bias and weak instrument issues are unlikely to affect the IV estimates.

The IV results of Table 1 indicate that an increase of the loan-to-income ratio by 1 standard deviation (+1.28) lowers the probability of approval by 17 to 22 percentage points. IV results thus suggest that the magnitude of non-instrumented estimates is substantially downward biased. This is consistent with a case where more creditworthy households tend to choose higher loan-to-income ratios. In such a scenario, unobservable characteristics of the applicant are positively correlated with the loan-to-income ratio and have a positive impact on the approval rate, leading to an upward bias on the coefficient.

Such approval regression results (Columns of coefficients) are then used to predict the probability of approval for all households in each census block group. In this out-of-sample estimation, the race of the household comes from the Census, and the loan-to-income ratio is computed using FNC housing prices and the total household income of the Census. Predictions are made under an assumption of a loan-to-value of 0.8, that is, the norm for conforming loans used for home purchase (Adelino, Gerardi & Willen 2013).[31]

4.3 Estimated Preferences and Willingness to Pay for Amenities

This section presents estimated preferences with and without approval origination constraints. Estimated household preferences are the outcome of the optimization of the moment conditions (Equation 23). Recall that preferences are allowed to vary across neighborhoods (this utility component

[30] The corresponding uninstrumented probit results are presented in Appendix Table B.
[31] Results are robust to the inclusion or exclusion of loans above the conforming loan limit.

defines the *base utility* δ), across individuals, and to vary across neighborhoods according to individual characteristics (Equation 1). Base utility depends on the neighborhood characteristics (Equation 2) which include: log median transaction price $log(p_j)$, the neighborhood's demographic composition (race, median income, and residents' education), school test scores (California's Academic Performance Index), and the quality of housing construction (age structure and number of rooms). As discussed in Section 3.1, the estimation of the relationship between base utility, log price, and neighborhood characteristics requires an IV approach as observable variables are likely to be correlated with unobserved determinants of neighborhood utility.

The instrumental variable strategy is as follows. First, the regression includes a location-specific blockgroup fixed effect, so that the impact of housing characteristics, local amenities, and neighbors' characteristics are estimated in changes across time rather than in cross section. Second, we use time-variation in the observable characteristics of geographically adjacent block groups as instruments for the decennial change in log price. Such instrumental variable strategy, in the spirit of Berry et al. (1995), is based on the assumption that neighboring blocks' changes in observable characteristics are less confounded by changes in unobservables than the block's own characteristics. Neighbors also sort endogenously in unobservable and observable dimensions across neighborhoods. We instrument the demographic composition (% black, % Hispanic, % Asian, log(median household income), % college educated) by neighboring tracts' demographic composition as in Bayer & Timmins (2007).

Selection of neighborhood characteristics of interest for inclusion in the utility regressions (1) and (2) builds on existing literature. The racial composition of neighborhoods and its interaction with household race allows to test for heterogenous racial preferences (Bayer et al. 2004, Bayer et al. 2007, Becker & Murphy 2009). The emphasis on the age of dwellings and its interaction with income is stressed in filtering theory (Rosenthal 2014). The inclusion of the number of rooms, as a proxy for the size of the dwellings, and its interaction with the distance to the metropolitan area's Central Business District, relates to the monocentric model, in which households sort at a particular distance from the CBD according to the relative importance of their taste for either larger houses, or for closer proximity with the city center and shorter commutes (Fujita 1989). The preference for more educated neighborhoods, in addition to more affluent neighbors, is linked to the literature on parents' involvement in school and education neighborhood peer effects (Durlauf 2004, Goux & Maurin 2007). Several of the neighborhood characteristics are interacted with individual characteristics such as race and income, thus allowing for a multidimensional sorting.

Table 2 presents estimation results. The first two columns present the estimated coefficient and the willingness to pay according to the model with borrowing constraints. The next two columns present the corresponding estimates for the model without borrowing constraints. In that later model, the probability that any neighborhood j belongs to the choice set of any household i in year t is set equal to one. The willingness to pay is the change in price that offsets a given change in the quality of amenities. Formally for a given amenity z_{jtk}, the willingness to pay $\Delta \log(p)$ for a change in amenity z_{jtk} is such that $-\alpha\Delta \log(p) = \beta \cdot \Delta(z_k)$, i.e.$-(\beta/\alpha) \cdot \Delta(z_k)$. In Table (2), we report the willingness to pay corresponding to a one percentage point increase for the variables expressed in fraction (fraction of each race, fraction of college graduates), and an increase of 10% of the standard deviation for the all other variables.[32]

Estimates of the preferences for neighborhoods' racial composition reveal sharp differences between the models with and without borrowing constraints. A non-black household's willingness to pay for a neighborhood with 1 percentage point fewer black neighbors is \$32,871 in the model with borrowing constraints vs. \$25,647 in the model without borrowing constraints. Black households have a preference for black neighbors, as their willingness to pay for a corresponding one percentage point increase in the fraction black is about \$10,555.[33] Non-black households' preferences for black neighbors are statistically different at 1% between the model with credit and the model without credit constraints; but the difference is not significant across the two models when it comes to black households' preferences for black neighbors.

Similar patterns of racial preferences arise when considering households' preferences for the fraction of Hispanics and Asians in a neighborhood. Hispanic and Asian households derive a positive utility from living in neighborhoods with more same-race neighbors.

The existence of approval constraints mitigates how preferences for neighborhoods' racial composition gets reflected into demand differences. As mortgage credit is more constrained in neighbor-

[32] As the distribution of the fraction black, Hispanic, or asian households across neighborhoods is substantially right-skewed (few neighborhoods with the majority of such residents), a one standard deviation change is a large percentile increase. We choose to display a 1 percentage point increase in this case. All other amenities, including the standardized Academic Performance Index, have symmetric distributions where a 1 standard deviation increases corresponds to an increase of about 60 to 70 percentile ranks.

[33] The table gives black households' willingness to pay of the fraction black in the neighborhood next to the coefficient of the interaction term between the fraction black and the black indicator variable. Similarly for Hispanic (resp. asian) households' willingness to pay for the fraction Hispanic (resp. asian) in the neighborhood.

hoods with a higher fraction of non-black households, the model with approval constraints reveals sharper underlying differences in preferences for non-black neighborhoods. Overall the model without borrowing constraints tends to underestimate the willingness-to-pay for neighborhoods with a higher non-black fraction because households are subject to high probabilities of denial in such neighborhoods.

The two models do not significantly differ with respect to the preference for housing characteristics, e.g. the structure's median age, the number of rooms or the distance to the Central Business District. However, the model estimated with borrowing constraints exhibits a significantly larger willingness to pay for neighborhoods with better performing public schools.[34] The willingness to pay for a 10% S.D. increase in the Academic Performance Index is \$3,265 for the average household in the model with borrowing constraints, and \$1,217 in the model without borrowing constraints. Willingness to pay for public school performance is higher for lower income levels: a household with a log income in the first quartile has a willingness to pay of \$9,312 for a 10% increase in the S.D. of test scores, in the model with borrowing constraints, as compared to \$4,964 in the model without borrowing constraints. Such difference is statistically significant at 1%.

Finally, the point estimate of the sensitivity of the base utility to a change in the median price of a given neighborhood is slightly stronger in the model without than in the model with borrowing constraints (-1.742 vs. -1.584) but the two coefficients are not statistically different. However as we shall see next, the price elasticity of the two model will be markedly different between the two model once the effect of price change on the probability of acceptance is taken into account.

4.4 Price Elasticity

The model with borrowing constraint allows separate estimation of two additive components of the price elasticity of demand. By price elasticity of demand, we refer here to the elasticity of the demand for a neighborhood with respect to its *own pric*e but we shall emphasize that the model provides estimates for the complete set of cross-price and income elasticities.[35] The first component, the conditional elasticity of demand, is computed assuming that approval-based choice sets are not

[34] The coefficient on API is stronger for households with income below the neighborhood mean income, a finding which is likely to capture the preference for private school education, in some neighborhoods, for households with an income high enough to afford it.

[35] The full matrix of total demand derivatives is used in the general equilibrium analysis of Section 5.

affected by a change in house prices. Absent heterogeneity, the conditional elasticity is simply computed as one minus the market share of a given neighborhood times the price-sensitivity of the base utility, and given the large number of neighborhoods (4,416), can be simply approximated by the price sensitivity. However, neighborhood utility includes a household-specific random coefficient on neighborhood log(price). This implies that conditional price elasticity will be heterogenous. The second component, the borrowing elasticity, captures the fall in demand associated with a fall in approval rates following an increase in housing prices. This elasticity is also heterogenous reflecting differences in the change in approval rates with respect to a change in prices across different neighborhoods.

The sum of the two components gives the total (unconditional) price elasticity. Figure 2 illustrates the elasticity decomposition with the distribution of elasticities across neighborhood (top panels of the figure) and their sensitivity to the initial log price (bottom panels of the same figure). Conditional demand elasticity is tightly centered around its mean (-1.58), while the borrowing elasticity is much more heterogeneous reflecting large differences in the approval response to a price increase across neighborhoods. Furthermore, there are striking differences in the relationship between initial housing price and each component of the price elasticity. The conditional elasticity is decreasing in the initial price level. At given borrowing constraints, demand for relatively expensive neighborhoods is less price-sensitive than the demand for relatively cheap neighborhoods. In other words, conditional elasticity is lower in magnitude for relatively expensive neighborhoods. The opposite is true for the borrowing elasticity. At a given income, buying in a more expensive neighborhood means a higher leverage and thus a lower approval probability. As a consequence of these two opposite effects, the total elasticity does not display any relationship with neighborhood prices. To evaluate this result, we can compare it to the demand elasticity in the model without borrowing constraints (Appendix Figure A). In this case, total demand elasticity is decreasing in housing prices and thus may not feature two important characteristics due to the mortgage approval channel: the high degree of heterogeneity in demand elasticity and its lack of relationship with house prices.

5 General Equilibrium Impact of Lending Standards

The model allows a simulation of a change in lending standards of the magnitude of the credit boom. The model predicts that such relaxation of lending standards typically leads to an increase in demand for most neighborhoods. This generally leads to excess demand that outstrips housing supply. The general equilibrium price change is the price change in each neighborhood that reduces excess demand to the point where each neighborhood's demand equals each neighborhood's housing supply.

Even though neighborhood demand stays unchanged, neighborhood demand by race, ethnicity, and income changes. Thus the model predicts substantial reallocation of households across space. We present in this section first the model's predicted impact of lending standards on the distribution of prices across neighborhoods, and second the model's predicted impact of lending standards on households' spatial segregation.

5.1 Comparative Statics: Price Impacts of Lending Standards

This section derives the impact of lending standards $(\boldsymbol{\psi}, \boldsymbol{\gamma}, \Psi)$ on neighborhood log prices $\log(\mathbf{p})$. We start by focusing on one specific coefficient, say, the coefficient of the log loan-to-income ratio in the approval specification, noted as a scalar ψ. A change in the coefficient ψ of the approval specification causes three different demand effects: first, a change in neighborhood demand at given prices. Second, the change in lending standards causes a change in each neighborhood's price (*the general equilibrium change in prices)* that in turn causes a change in demand. Third, the change in lending standards causes a change in neighborhoods' racial composition, which affects households' preferences for these specific neighborhoods. Such three effects formally translate into:

$$\frac{d\mathbf{D}}{d\log(\mathbf{p})} = \underbrace{\frac{\partial \mathbf{D}}{\partial \psi}}_{Direct\,Impact} + \underbrace{\frac{\partial \mathbf{D}}{\partial \log(\mathbf{p})} \cdot \frac{d\log(\mathbf{p})}{d\psi}}_{Impact\,due\,to\,price\,changes} + \underbrace{\frac{\partial \mathbf{D}}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \psi}}_{Impact\,due\,to\,social\,preferences} \tag{21}$$

For the sake of clarity, we consider here the case of perfectly inelastic housing supply ($\frac{\partial \mathbf{S}}{\partial \log(\mathbf{p})} = 0$), and extend the analysis to the non-perfectly inelastic housing supply case in Section 6.1.[36] The

[36] The evidence presented in Section 6.1 suggests indeed that housing supply is fairly inelastic in most neighborhoods of San Francisco. Data from Saiz (2010) suggests that the metropolitan-area wide housing supply elasticity is 0.66 for the entire San Francisco metropolitan area and 0.76 for the entire San Jose metropolitan area.

price change that maintains demand equal to supply in each neighborhood is such that:

$$d\log(\mathbf{p})/d\psi = \left[\frac{\partial}{\partial \log(\mathbf{p})}\mathbf{D}(\log(\mathbf{p}),\psi) - \frac{\partial \mathbf{S}}{\partial \log(\mathbf{p})}\right]^{-1}\left[-\frac{\partial}{\partial \psi}\mathbf{D}(\log(\mathbf{p}),\psi) - \frac{\partial \mathbf{D}}{\partial \mathbf{v}}\cdot\frac{\partial \mathbf{v}}{\partial \psi}\right] \tag{22}$$

a vector of size $\sum_t J_t$, i.e. one vector of log price changes per year of analysis. The vector of prices $\log(\mathbf{p})$ affects demand through its impact on the base utility vector $\boldsymbol{\delta}$.

(i) Impact of Lending Standards on Demand, at Given Prices

We start with the impact of lending standards on demand at given prices. Lending standards affect both the probability of each choice set, and the choice of the household within the choice set. We start by focusing on the impact of lending standards on the probability of the choice set.

$$\frac{\partial}{\partial \psi}D(j,t|\boldsymbol{\delta},\mathbf{z}_{\cdot t};\psi) = \sum_{C\in\mathbf{C}}\int_X \left[\frac{\partial}{\partial \psi}P(C|\mathbf{z}_{\cdot t},\mathbf{x}_{it};\psi)\right]\cdot P(j,t|\boldsymbol{\delta},\mathbf{z}_{\cdot t},\mathbf{x}_{it},C_{it})\cdot f(\mathbf{x}_{it})d\mathbf{x}_{it} \tag{23}$$

Empirically the sum $\sum_{C\in\mathbb{C}}$ is taken over the set of simulated choice sets. The derivative of the probability of the choice set $P(C|\mathbf{z}_{\cdot t},\mathbf{x}_{it};\psi)$ is the derivative of a product of J terms $\Pi_{j\in C}\phi_{jt}\Pi_{k\notin C}(1-\phi_{kt})$ w.r.t. lending standards. Each individual probability ϕ_{jt} has a simple binomial derivative, and thus the impact of lending standards on choice set probabilities:

$$\frac{\partial P(C|\mathbf{z}_{\cdot t},\mathbf{x}_{it},\psi)}{\partial \psi} = P(C|\mathbf{z}_{\cdot t},\mathbf{x}_{it},\psi)\cdot\underbrace{\sum_j w_{ijt}\cdot\{\mathbf{1}(j\in C)-\phi_{jt}(\mathbf{z}_{jt},\mathbf{x}_{it};\psi)\}}_{\Upsilon_{ijt}(C)} \tag{24}$$

where w_{ijt} is the value corresponding to the coefficient ψ. The impact of lending standards on the probability of approval in any one neighborhood (a single term in the sum Σ_j) is of the sign of w_{ijt}. For instance, if w_{ijt} corresponds to the loan-to-income ratio, w_{ijt} is positive, the coefficient ψ of the loan-to-income ratio in the approval specification is negative. An increase in ψ will lead to an increase in the probability ϕ_{ijt} for all households i and all neighborhoods j of the city. And the probability $P(C)$ of all choice sets C increases. The Monte Carlo simulated estimate of (23) is obtained by averaging (23) over simulated choice sets.

(ii) Impact of Prices on Demand, at Given Lending Standards

We then turn to the impact of log prices on demand. The matrix of demand derivatives $\frac{\partial}{\partial \log(\mathbf{p})}\mathbf{D}(\log(\mathbf{p}), \psi)$ is considered for year t. $\log(\mathbf{p})$ is a line vector of size J_t while $\mathbf{D}(\mathbf{p}, \psi)$ is a column vector of size J_t. We distinguish diagonal elements, which are derivatives of demand w.r.t.[37] its own price (proportional to own-price demand elasticity), and off-diagonal elements, which are derivatives of demand w.r.t. another neighborhood's price (proportional to cross-price demand elasticity).

The derivation of the cross-price demand elasticity yields a decomposition between borrowing elasticity and conditional price elasticity, which is the analog to the one presented in 2 for the own-price demand elasticity:

$$\begin{aligned}\frac{\partial D(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}; \psi)}{\partial \log(p_k)} &= \underbrace{\sum_{C\in\mathbf{C}} \int_X \frac{\partial}{\partial \log(p_k)} P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi) \cdot P(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it}}_{\text{Shift in choice set}} \\ &+ \underbrace{\sum_{C\in\mathbf{C}} \int_X P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi) \cdot \frac{\partial}{\partial \log(p_k)} P(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it}}_{\text{Shift in utility}} \quad (25)\end{aligned}$$

Appendix Section C provides closed-form derivations for both demand derivative terms.

(iii) Impact of Lending Standards on Social Preferences

We derive the impact of lending standards on neighborhood composition. Neighborhood composition affects neighborhood demand as long as households have preferences for particular neighbor demographics (e.g. education, race, ethnicity). Getting such social interaction effects is the last step in completing the calculation of general equilibrium effects on prices. Formally, such social interaction effect is a first-order impact on households' demand $-\frac{\partial \mathbf{D}}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \psi}$. The second factor, the impact of lending standards on neighborhood demographics, $\frac{\partial \mathbf{v}}{\partial \psi}$ is simply the vector of demands scaled by the total demand for the neighborhood.

$$\frac{\partial \mathbf{v}}{\partial \psi} = \left(\left(n^k/n\right) \cdot \left(\frac{\partial \mathbf{D}^k}{\partial \psi}/\mathbf{D}\right)\right)_{k=1,2,\ldots,M}$$

[37] We note "with respect to" as w.r.t. in the remaining parts of this paper.

where n^k/n is the fraction of population k in the overall population of the metropolitan area. The first factor is the impact of neighborhood demographics on demand. $\frac{\partial \mathbf{D}}{\partial \mathbf{v}}$ is also straightforwardly derived. For a group k, say Hispanics, $\gamma^k(\mathbf{x}_{it})$ is the coefficient of social interactions for an individual with characteristics $\mathbf{x}_{it}$. $\gamma^k(\mathbf{x}_{it})$ depends on $\mathbf{x}_{it}$ as Asians, Blacks, Hispanics, Whites do not necessarily exhibit similar preferences. Then:

$$\begin{aligned}\frac{\partial}{\partial \mathbf{v}^k} D(j,t|\boldsymbol{\delta}_{\cdot t},\mathbf{z}_{\cdot t};\psi) &= \sum_{C\in\mathbf{C}}\int_X P(C|\mathbf{z}_{\cdot t},\mathbf{x}_{it};\psi) \\ &\quad \cdot\gamma^k(\mathbf{x}_{it})\cdot P(j,t|\boldsymbol{\delta}_{\cdot t},\mathbf{z}_{\cdot t},\mathbf{x}_{it},C_{it})\cdot(1-P(j,t|\boldsymbol{\delta}_{\cdot t},\mathbf{z}_{\cdot t},\mathbf{x}_{it},C_{it}))f(\mathbf{x}_{it})d\mathbf{x}_{it}\end{aligned}$$

$\gamma^k(\mathbf{x}_{it})$ varies across individuals: for instance, results suggest households' preferences for same-race neighbors.

(iv) Empirical Results: General Equilibrium Impacts of Lending Standards on Prices

The general equilibrium impact of lending standards on prices is the shift in prices that keeps demand for each neighborhood constant. We combine the analytic formulas of (i)–(iii) above to derive the following: first, the impact of lending standards on demand at constant prices; second, the impact of prices on demand; and third, the impact of lending standards on neighborhood racial and income composition. All three effects lead to the general equilibrium impact of lending standards on prices, following equation (22) above.

The derivations presented in the previous subsections (i)–(iii) were done by looking at the marginal change of one arbitrary coefficient of the approval specification. Table D estimates the change in lending standards between 2000 and 2006, and suggests that there were changes in underwriters' sensitivity to the log(price) of the house, the log(income) of the applicant, and to the race of the applicant. The regression is a panel logit regression combining data from the Home Mortgage Disclosure Act of 2000 and 2006. The specification is similar to the main baseline analysis of approval rates (Table 1), with added interactions between the 2006 year dummy and respectively: the log(price), the log(income) of the applicant, and the applicant's race. Results suggest a relaxation of lending standards in each of those dimensions. We combine the comparative statics in each of these dimensions to obtain a marginal change in lending standards comparable to the 2000–2006 relaxation.

Figure 3 illustrates the effect of such change in lending standards on mortgage approval, housing demand, and the distribution of housing prices. Figure 3 (d) plots the average marginal impact of such relaxation on the individual probability of approval. Almost all neighborhoods experience an increase in average approval probability, except for the neighborhoods with the highest prices (log(prices) above 14, i.e. prices above \$1.2M). The rise in the approval probability is higher in neighborhoods with lower prices. Figure 3 (c) plots the implied partial equilibrium effect, where the increase in demand (vertical axis) at constant prices is plotted against the initial log(price) in 2000. For neighborhoods with the lowest initial price, demand increases by more than 20% on average, while neighborhoods with the highest initial price see a decline in demand at partial equilibrium. Figure 3(f) shows that neighborhoods with low initial prices are also neighborhoods with a high initial fraction black. Black households' preferences for same-race neighbors thus leads to an increase in black demand for black neighborhoods, an effect depicted in Figure 3 (e) ("social interaction impacts", noted $\partial \mathbf{D}/\partial \mathbf{v} \cdot \partial \mathbf{v}/\partial \psi$).

Figure 2(f) showed that there is little correlation between the initial price level and total demand elasticity. Hence the partial equilibrium effect of lending standards on demand Figure 3(c) combined with the social interaction impacts of Figure 3(e) explain why the general equilibrium impact of lending standards on prices leads to a greater increase in prices in neighborhoods with lower prices (Figure 3(a)). The general equilibrium impact on prices is obtained following equation 22, and demeaned. As the change in lending standards is calculated over 2000-2006, the vertical axis of Figure 3(a) is the annual average of log(prices).

House Price Changes: Model vs. Data

Next, we compare the prediction of the model on log(price) changes to actual price changes during the same the period computed from property transaction data. While the *change* in lending standards is taken as realized in the data, the estimation of households' preferences for neighborhoods, and estimation of initial lending standards did not use annual information for the boom years of 2001-2006. The last observation in our decennial panel is 2010; in 2010 the Case-Shiller price index for the San Francisco MSA was almost back to its 2000 level (136.99 vs. 130.06).

Figure 4 plots the actual log price change in the model (red dots) and in the data (black dot) as a function of the initial log price in 2000. As in Figure 3 we use the average annual log(price) change over 2000-2006. Figure 4 indicates that the model reproduces the compression of the price distribution, i.e. the negative correlation between neighborhood price increases and the initial price level. Such correlation has been observed in this paper in the San Francisco Bay area, and in San Diego by Landvoigt, Piazzesi & Schneider (2015). The model predicts a correlation of -0.435 (bottom table of the Figure), while observations of actual log(price) changes imply a correlation of -0.630. The model's predictions also match key moments of the data. Results reveal that the model matches the annualized price increase well (+0.125 vs +0.111); predictions of log price changes explain about 46% of the standard deviation of the realizations. The median of predictions is also similar to the median in the data (0.113 vs. 0.109) and the upper quantiles are also close (0.131 vs 0.149).

Finally, we compare, at the neighborhood-level, predicted price changes to actual price changes by regressing the actual neighborhood log(price) change on the predicted neighborhood log(price) change (Figure 5). The blockgroup-level prediction and the actual log(price) change are significantly positively correlated at 1%, with a slope of 0.62 and a correlation of 0.29. The variance of predicted log(price) changes explains 8.3% of the variance of actual log(price) changes. Even if the fundamental shock we used for comparative statics is the actual relaxation of lending standards, this is still a rather notable model performance given the notorious difficulty of predicting house price changes.

5.2 Impacts of Lending Standards on Spatial Segregation

The general equilibrium price change maintains demand in each neighborhood constant. However, the demand of particular racial or other demographic subgroups for each neighborhood is typically affected by changes in lending standards.

Consider, for example, the demand D_{rjt} of racial group $r \in \{white, hispanic, black, asian, other\}$ for a specific neighborhood j in year t. The total change in demand caused by a change in lending standards is decomposed into a partial equilibrium effect, an effect due to social preferences, and a change due to shifts in the distribution of prices:

$$\frac{dD_{rjt}}{d\psi} = \underbrace{\frac{\partial D_{rjt}}{\partial \psi}}_{partial\,equilibrium} + \underbrace{\frac{\partial D_{rjt}}{\partial \log(\mathbf{p})} \cdot \frac{d\log(\mathbf{p})}{d\psi}}_{general\,equilibrium} + \underbrace{\frac{\partial D_{rjt}}{\partial \mathbf{v}} \cdot \frac{\partial \mathbf{v}}{\partial \psi}}_{preferences\,for\,same\,race\,neighbor} \tag{26}$$

and in general $dD_{rjt}/d\psi \neq 0$, while $d\mathbf{D}/d\psi = 0$. Given racial groups' changing demand, the model gives a structural estimate of how changes in lending standards affect racial segregation across neighborhoods within the metro area.

As measure of Bay Area-wide racial segregation we use the exposure indices, as in Massey & Denton (1988) and Cutler, Glaeser & Vigdor (1999). The exposure index of a given racial subgroup (say, Blacks) to another (say, Whites) measures the average fraction of white neighbors for an average black resident. Thus:

$$Exposure(Whites|Blacks)_t = \sum_{j=1}^{J} \frac{Black_{jt}}{Black} \cdot \frac{White_{jt}}{Population_{jt}} \tag{27}$$

where the first factor $Black_{jt}/Black$ is the share of black population living in neighborhood j in year t; and the second factor $White_{jt}/Population_{jt}$ is the fraction white in neighborhood j in year t. The isolation of Blacks is the exposure of Blacks to their own racial group. Therefore an increase in black isolation is due to a decline in black exposure to other races.

Definition (27) can be simply tied to the structural model: black population $Black_{jt}$ is simply the total black population count in the metro area multiplied by black neighborhood demand for that area, i.e. $Black_{jt} = Black \cdot D_{black,jt}$, as $D_{black,jt} \in (0, 1)$. Similarly, $White_{jt} = White \cdot D_{white,jt}$.

Thus the model provides a structural prediction of the impact of lending standards on racial isolation and exposure. For instance, deriving black exposure to Whites with respect to the lending standard parameter ψ:

$$\frac{d}{d\psi} Exposure(Whites|Blacks)_t = \frac{White}{Population} \cdot \frac{d}{d\psi} \left[\sum_{j=1}^{J} \frac{D_{black,jt} \cdot D_{white,jt}}{D_{jt}} \right] \tag{28}$$

The change in exposure is driven by both white and black demand changes.

The resultant analysis of the impact of lending standards on neighborhoods' racial composition and on racial segregation are presented in Figure 6 and in Table 3. The starting point of such comparative statics is the spatial distribution of racial groups in the Bay area in 2000.

Figure 6(a) is a scatter plot where each dot is a neighborhood, and the vertical axis is the change in white demand for each neighborhood, while the horizontal axis is the fraction black in 2000. Black dots are demand changes *at given prices*, as a percentage of initial demand. The red dots are the neighborhoods' predicted change in white demand *accounting for the general equilibrium* change in prices. The figure suggests that a relaxation of lending standards leads to a substantially higher increase in white demand in neighborhoods that are majority non-black than in neighborhoods with more than 10 to 20% black population. Interestingly, as prices respond to the more lenient lending standards, white population change in majority non-black neighborhoods is less than the white population change predicted by the partial equilibrium changes, but remains substantially higher in majority non-black neighborhoods. Thus Figure 6(a) suggests that the exposure of Blacks to White declines when lending standards are relaxed.

Figure 6(b) (resp., (d)) suggests a similar pattern for Hispanics (resp., Asians): the relaxation of lending standards should lead to a decline in black exposure to Hispanics (resp. to Asians).

Interestingly, Figure 6(c) shows that black mobility to majority black neighborhoods is muted by general equilibrium price changes. Indeed, at given prices (black dots), the relaxation of lending standards would lead to substantially higher black demand for black neighborhoods than for majority non-black neighborhoods. But, when accounting for the change in prices at equilibrium (red dots), the rise in black demand does not exceed 4-5% of initial 2000 demand.

Figure 6(e) suggests that there is a broadly positive correlation between the fraction black in the neighborhood in 2000 and the increase in log price, as expected given Figure 3 and the negative correlation between the fraction black and the log(price). However, the relationship between the increase in log(price) and the initial fraction black, displayed in Figure 6(e), is non-monotonic and U-shaped: the largest price increases are observed in neighborhoods with either a low share of black population or with a high share of black population. The price increase in majority non-black neighborhoods is not high enough to offset the partial equilibrium rise in white demand for non-black neighborhoods.

Such neighborhood-level analysis is summarized in Table 3, which presents the comparative statics impact of lending standards on our racial segregation measures. Table (a) presents such impacts at given prices, while Table (b) presents the total impact, once price changes are accounted for. The diagonal of each table is the isolation of Whites, Blacks, Hispanics, Asians (the exposure of a racial group to itself), while off-diagonal elements present the impact on the exposure of a racial group to another. All estimates are in percentage points, as isolation and exposure measures are expressed in percentages.

Results suggest that a relaxation of lending standards will lead to a decline in the exposure of Blacks to Whites, by 2.04 percentage points, and a decline in the Exposure of Whites to Blacks, by 0.36 percentage points. This decline is partially offset by the increased exposure to Hispanics and Asians. This decline is not predicted by partial equilibrium analysis: Table 3(a) shows that a partial equilibrium analysis would lead to predict an increase in black exposure to Whites and an increase in white exposure to Blacks.

Such segregation results, provided by the comparative statics of the estimated model are in line with more aggregate metropolitan-level empirical findings of Ouazad & Rancière (Forthcoming). When considering the sample of all U.S. metropolitan areas, the relaxation of lending standards during the credit boom is indeed associated with a reduction in the exposure of Whites to Blacks. The model goes beyond the metro-level results: first the estimated model provides a direct quantification of the population flows, e.g. shows that the decreased exposure is mostly driven by white mobility in to mostly white neighborhoods. Second, it provides a decomposition of the relative importance of general vs. partial equilibrium effects; finally it shows how the greater price increase in less expensive neighborhoods is consistent with a lower exposure of Blacks to Whites.

Beyond the broad picture of an increase in segregation, the results also suggest the possibility of limited white gentrification of a number of initially mostly black neighborhoods. We observe in Figure 6 (a) a modest increase in the demand of Whites for mostly black neighborhoods. In that case the partial equilibrium demand effect is not muted by general equilibrium effect despite the sharp increase in house prices in these neighborhoods (Figure 6 (b)). By contrast, in the very same neighborhoods, the increase in the partial equilibrium demand by Blacks is strongly muted in general equilibrium. Put together this evidence points toward some limited gentrification of the kind observed in some neighborhoods of Oakland during the credit boom.

6 Discussion

6.1 The Role of Housing Supply Elasticity

The general equilibrium analysis of Section 5 assumed that the supply of housing units was constant during the period. We relax here this assumption and check whether the compression of the price distribution, observed between 2000 and 2006, could be due to the heterogeneity in housing supply elasticities across neighborhoods rather than to the differential impact of lending standards changes.

We build neighborhood- (i.e. blockgroup-) level estimates of housing supply elasticities by combining satellite data on land cover and elevation in order to measure the share of each blockgroup that is not developed in 1992 and can likely be developed given the geographic features of the land. Then, in a similar fashion as in Saiz (2010), but in our case at the blockgroup level, we perform the following regression of log housing units on log price:

$$\begin{aligned} \log(housing\,units_{j,t}) &= a + (b + c \cdot Undeveloped\,Share_j + d \cdot Elevation_j) \cdot \log(price_{j,t}) \\ &\quad + Blockgroup_j + Residual_{j,t} \end{aligned} \tag{29}$$

where j is a neighborhood index, t is either 2000 or 2010; $Undeveloped\,Share_j$ is the undeveloped share of the blockgroup's surface in 1992, 8 years prior to the first data used in the regression. The measure of undeveloped share of land corresponds to the share of the surface of the blockgroup (in squared meters) that is not developed and that is not covered by water. The measure is derived from satellite data of the United States Geological Survey (U.S.G.S.) database. The landcover data set of 1992 measures, in each 30m×30m cell, whether the cell is developed (low or high intensity, residential or commercial) and the nature of the land (forest, water, barren, rock, grass, wetland, crops and pasture). Appendix Figure B (a) illustrates the construction of our underdeveloped share measure by mapping the underlying satellite landcover data in the city of San Jose.

Then we use a set of set of elevation measures ($Elevation_j$) based on the slope and the ruggedness of the blockgroup to proxy for the cost of building on undeveloped cells.[38] Such elevation measures come from a second satellite data set, the U.S.G.S.'s digital elevation model (D.E.M.), which measures elevation for each 30m×30m cell.

$Blockgroup_j$ is a blockgroup fixed effect. Neighborhood-level housing supply elasticity is proxied by the coefficient $\eta_j = b + c \cdot Undeveloped\,Share_j + d \cdot Elevation_j$ in specification 29 above. Appendix Figure C(a) presents the distribution of estimated supply elasticity across neighborhoods. About half of the neighborhoods exhibit a supply elasticity inferior to 0.1, and about 82% percent an elasticity below 0.2. As expected neighborhood housing supply elasticity is positively correlated

[38] The ruggedness index measures terrain irregularity and was initially developed by Riley, DeGloria & Elliot (1999) and used in economics by Nunn & Puga (2012). The ruggedness at a given cell (30mx30m) is the square root of the sum of the squared differences of the elevation of the cell with its eight adjacent cells.

with the distance to the Central Business District (Appendix Figure B (b)).

We then replace each neighborhood housing supply change $dS_{jt}/d\log(p_j) = S_{jt}\eta_j$ by its empirical counterpart in the general equilibrium price change equation 22. This allows a robustness check for the estimates of Section **5**, i.e. the robustness of estimated general equilibrium price effects to the introduction of such neighborhood-level housing supply elasticity (η_j).

Appendix Figure C(b) compares the price change with perfectly inelastic supply (black points) and with such local, neighborhood-level supply elasticity (red circles). The black (resp. red) line is the OLS regression of log(price) change on initial log price level in 2000. Results presented in this figure suggest little impact of those levels of housing supply elasticity on estimated general equilibrium impacts of lending standards on prices.

6.2 Tenure Choice: Homeownership and Rental

The relaxation of lending standards may affect both neighborhood choice and tenure choice. Changes in lending standards, by affecting the probabilities to obtain credit, are also likely to change the relative demand for ownership vs. rental housing both within and across neighborhoods. By introducing renting in our framework, we can discuss whether this relative demand effect has a significant impact on the results of Section 5.

We introduce tenure choice in the model as follows. In each neighborhood j, household i can choose either homeownership $s = ownership$ or rental, $s = rental$. The utility of neighborhood j with tenure s is noted U_{ijst}, as in equation 1, and the corresponding base utility δ_{jst}. The difference $U_{ij,ownership,t} - U_{ij,rental,t}$ is partly driven by unobservable quality differences between the rental housing stock and the owner-occupied housing stock; and driven by households' preference for homeownership vs. rental. Such model with neighborhood and tenure choice reduces to the paper's baseline model of Section 2 whenever (i) a uniform arbitrage relationship ties the price of owner-occupied units and rental values, such as the the price of owner-occupied units equal to the discounted value of rental payments with a constant discount factor; and (ii) households are indifferent between homeownership and rental.

The model with neighborhood and tenure choice relaxes those two assumptions. Rental prices are assumed to be linked to the price of owner-occupied housing. However, by estimating a different base utility for rental and homeownership in the same neighborhood, the model allows for a neighborhood-and time-specific relationship between rental prices and homeownership prices.

We then estimate the model with neighborhood choice and tenure choice, and the associated general equilibrium impact on the price of owner-occupied units. Results, presented in Appendix Figure D, suggest no economically significant difference between the two models in the predicted magnitude of the compression of the price distribution. Both models predict a similar (and not statistically different) correlation between log price changes and the initial log price in 2000.

7 Conclusion

This paper's objective was to put forward and estimate a model of neighborhood choice that explicitly takes into account the role of borrowing constraints. The new methodology developed is based on the idea that differences in probability of mortgage approval across neighborhoods shape the choice set of prospective home buyers. This approach yields results that are substantially different from those obtained from models omitting the role of mortgage credit.

We show and explain how key preference features such as the additional premium put to live in neighborhoods with high quality schools, or the preference of Whites to live among racial peers, can be substantially underestimated when omitting the role of borrowing constraints. The model with borrowing constraints also allows separate estimation of two key components of the price-elasticity of demand, one which reflects preference-based willingness to pay, and the other who reflects the impact of price changes on the probability to obtain a mortgage. These new elasticity estimates could be provide a useful guidance to (i) policymakers estimating the impact of taxation on prices, and to (ii) potential real estate developers seeking to price new housing units.

The estimated model enables to run a large variety of counterfactual experiments and to distinguish, in each case, the role played by partial vs. general equilibrium effects, preferences vs. approval constraints. The model also accounts for social multiplier effects. In this paper, we focus on the effect of a variation in lending standards on the dynamics of house prices and segregation. We show that the model makes two out-of-sample predictions that are well in line with the data: (i) the compression in the dispersion of prices, and (ii) the increase in segregation.

There are many other dimensions in which the estimated model could be used to understand the dynamics of cities. For example, it could be used to assess the effects of a change in the level or distribution of income, of a change in the distribution of the quality of schools, or in the supply of transportation. The model could also be used “in reverse” in order to trace down the fundamental causes of salient city transformations such as the gentrification of some neighborhoods.

References

Adelino, M., Gerardi, K. & Willen, P. S. (2013), ‘Why don’t lenders renegotiate more home mortgages? Redefaults, self-cures and securitization’, *Journal of Monetary Economics* **60**(7), 835–853.

Adelino, M., Schoar, A. & Severino, F. (2012), Credit supply and house prices: evidence from mortgage market segmentation, Technical report, National Bureau of Economic Research.

Alonso, W. et al. (1964), *Location and land use: toward a general theory of land rent.*, Cambridge, Mass.: Harvard Univ. Pr.

Bajari, P. & Kahn, M. E. (2012), ‘Estimating housing demand with an application to explaining racial segregation in cities’, *Journal of business & economic statistics* .

Bayer, P., Ferreira, F. & McMillan, R. (2007), ‘A unified framework for measuring preferences for schools and neighborhoods’, *Journal of Political Economy* .

Bayer, P., McMillan, R. & Rueben, K. S. (2004), ‘What drives racial segregation? new evidence using census microdata’, *Journal of Urban Economics* **56**, 514–535.

Bayer, P. & Timmins, C. (2007), ‘Estimating equilibrium models of sorting across locations’, *The Economic Journal* **117**(518), 353–374.

Becker, G. S. & Murphy, K. M. (2009), *Social economics: Market behavior in a social environment*, Harvard University Press.

Berry, S., Levinsohn, J. & Pakes, A. (1995), ‘Automobile prices in market equilibrium’, *Econometrica* **63**(4), 841–90.

Conlon, C. T. & Mortimer, J. H. (2013), ‘Demand estimation under incomplete product availability’, *American Economic Journal: Microeconomics* **5**(4), 1–30.

Cutler, D. M., Glaeser, E. L. & Vigdor, J. L. (1999), 'The rise and decline of the american ghetto', *Journal of Political Economy* **107**(3), 1–52.

Debreu, G. (1970), 'Economies with a finite set of equilibria', *Econometrica: Journal of the Econometric Society* pp. 387–392.

Dell'Arriccia, G., Igan, D. & Laeven, L. (2009), 'Credit booms and lending standards: Evidence from the subprime mortgage market', *European Banking Center Working Paper* pp. 1–44.

Dubé, J.-P., Fox, J. T. & Su, C.-L. (2012), 'Improving the numerical performance of static and dynamic aggregate discrete choice random coefficients demand estimation', *Econometrica* **80**(5), 2231–2267.

Durlauf, S. N. (2004), 'Neighborhood effects', *Handbook of regional and urban economics* **4**, 2173–2242.

Fan, Y. (2013), 'Ownership consolidation and product characteristics: A study of the us daily newspaper market', *The American Economic Review* **103**(5), 1598–1628.

Favara, G. & Imbs, J. (2015), 'Credit supply and the price of housing', *The American Economic Review* **105**(3), 958–992.

Fujita, M. (1989), *Urban economic theory: land use and city size*, Cambridge university press.

Geweke, J. (1988), 'Antithetic acceleration of monte carlo integration in bayesian inference', *Journal of Econometrics* **38**(1), 73–89.

Geweke, J. (1996), 'Monte carlo simulation and numerical integration', *Handbook of computational economics* **1**, 731–800.

Glaeser, E. L., Gottlieb, J. D. & Gyourko, J. (2012), Can cheap credit explain the housing boom?, *in* 'Housing and the Financial Crisis', University of Chicago Press, pp. 301–359.

Glaeser, E. L., Gottlieb, J. D. & Tobio, K. (2012), 'Housing booms and city centers', *The American Economic Review* **102**(3), 127.

Goeree, M. S. (2008), 'Limited information and advertising in the us personal computer industry', *Econometrica* **76**(5), 1017–1074.

Goux, D. & Maurin, E. (2007), 'Close neighbours matter: Neighbourhood effects on early performance at school', *The Economic Journal* **117**, 1–23.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments estimators', *Econometrica: Journal of the Econometric Society* pp. 1029–1054.

Holmes, T. & Sieg, H. (2015), 'Structural estimation in urban economics', *Handbook of Urban and Regional Economics* **5**.

Imbens, G. W. & Lancaster, T. (1994), 'Combining micro and macro data in microeconometric models', *The Review of Economic Studies* **61**(4), 655–680.

Landvoigt, T., Piazzesi, M. & Schneider, M. (2015), 'The housing market(s) of San Diego', *American Economic Review* **105**(4), 1371–1407.

Loutskina, E. (2011), 'The role of securitization in bank liquidity and funding management', *Journal of Financial Economics* **100**(3), 663–684.

Loutskina, E. & Strahan, P. E. (2009), 'Securitization and the declining impact of bank finance on loan supply: Evidence from mortgage originations', *The Journal of Finance* **64**(2), 861–889.

Mas-Colell, A., Whinston, M. D., Green, J. R. et al. (1995), *Microeconomic theory*, Vol. 1, Oxford university press New York.

Massey, D. S. & Denton, N. A. (1988), 'The dimensions of residential segregation', *Social Forces* **67**(2), 281–315.

McFadden, D. (1973), 'Conditional logit analysis of qualitative choice behavior', *Frontiers in Econometrics* pp. 105–142.

Nevo, A. (2000), 'Mergers with differentiated products: The case of the ready-to-eat cereal industry', *The RAND Journal of Economics* pp. 395–421.

Newey, W. K. (1987), 'Efficient estimation of limited dependent variable models with endogenous explanatory variables', *Journal of Econometrics* **36**(3), 231–250.

Nunn, N. & Puga, D. (2012), 'Ruggedness: The blessing of bad geography in Africa', *Review of Economics and Statistics* **94**(1), 20–36.

Ondrich, J., Ross, S. & Yinger, J. (2003), 'Now you see it, now you don't: Why do real estate agents withhold available houses from black customers?', *Review of Economics and Statistics* **85**(4), 854–873.

Ouazad, A. & Rancière, R. (Forthcoming), 'Credit standards and segregation', *Review of Economics and Statistics* .

Petrin, A. (2002), 'Quantifying the benefits of new products: The case of the minivan', *Journal of Political Economy* **110**(4), 705–729.

Riley, S. J., DeGloria, S. D. & Elliot, R. (1999), 'A terrain ruggedness index that quantifies topographic heterogeneity', *Intermountain Journal of Sciences* **5**(1-4), 23–27.

Rosenthal, S. S. (2014), 'Are private markets and filtering a viable source of low-income housing? estimates from a "repeat income" model', *The American Economic Review* **104**(2), 687–706.

Saiz, A. (2010), 'The geographic determinants of housing supply', *The Quarterly Journal of Economics* pp. 1253–1296.

Su, C.-L. & Judd, K. L. (2012), 'Constrained optimization approaches to estimation of structural models', *Econometrica* **80**(5), 2213–2230.

Train, K. (2009), 'Discrete choice models', pp. 1–42.

Figure 1: Bank Branches and Liquidity

This map presents the set of bank branches from the Summary of Deposits in 1994. The map only includes bank branches whose corresponding bank is regulated by the Federal Reserve, the Office of the Comptroller of the Currency, or the Federal Deposit Insurance Corporation. The bank's liquidity level is derived from the Federal Reserve Bank of Chicago's Reports of Income and Condition (Call Reports). Color codes indicate the bank's liquidity level.

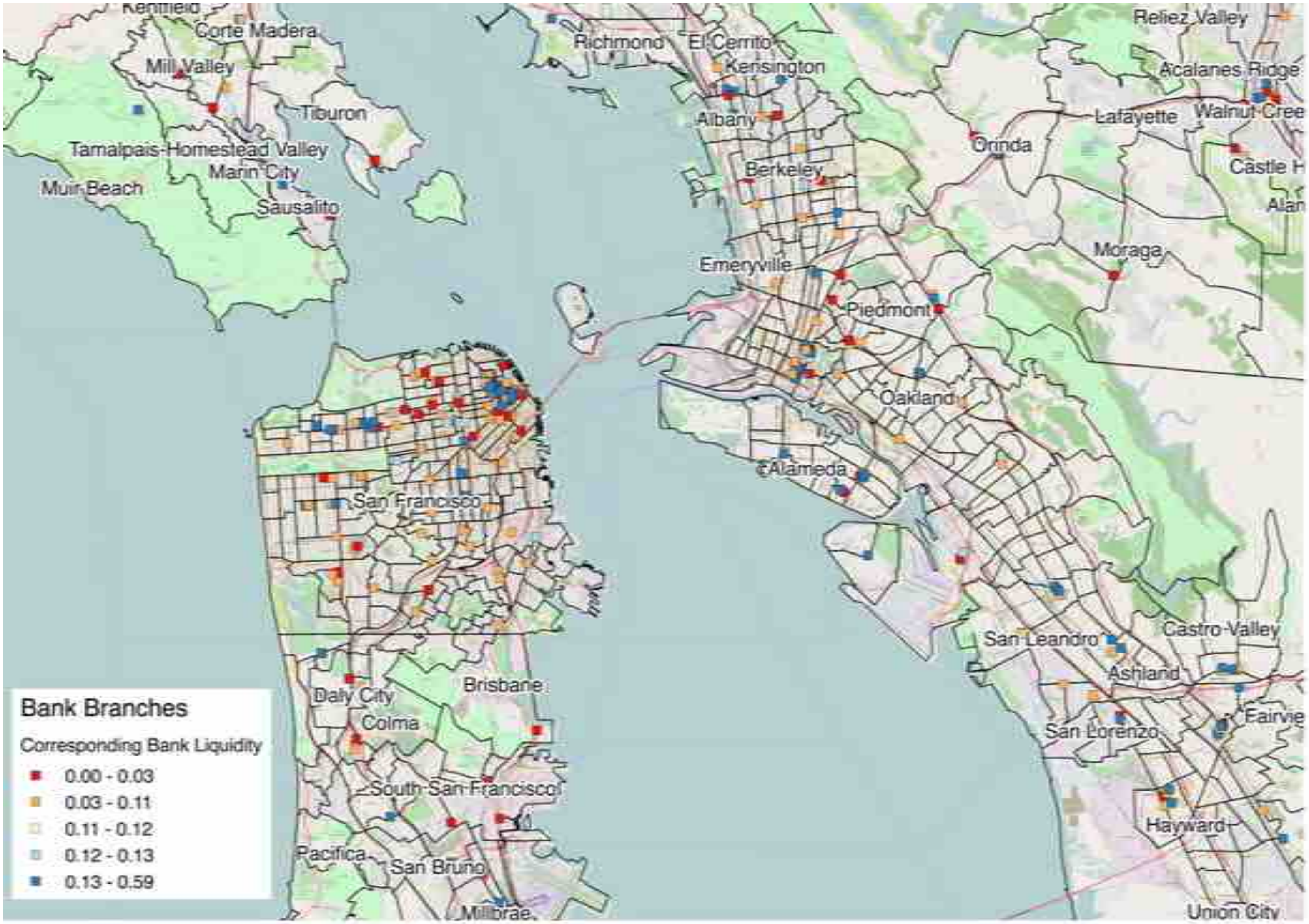
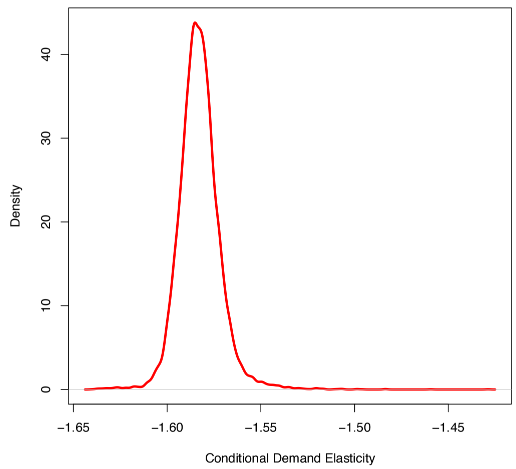


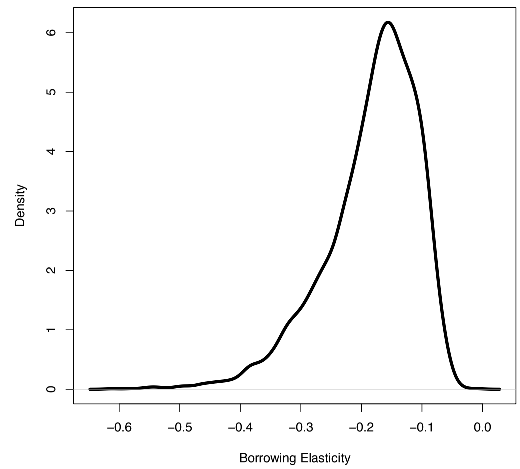
Figure 2: Demand Elasticity Distributions – Model with Borrowing Constraints

These figures present the total own-price demand elasticity η_j for each of the 4,416 neighborhoods, as well as the conditional own-price demand elasticity $\eta|_{C,j}$ and the borrowing elasticity $\eta_j - \eta|_{C,j}$. The conditional demand elasticity is the elasticity at given borrowing constraints, i.e. at given choice set.

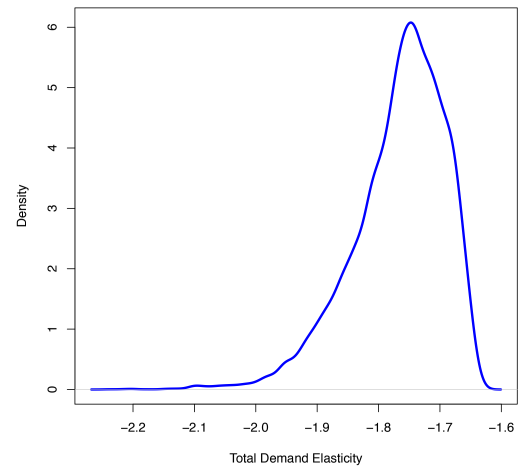
(a) Conditional Demand Elasticity



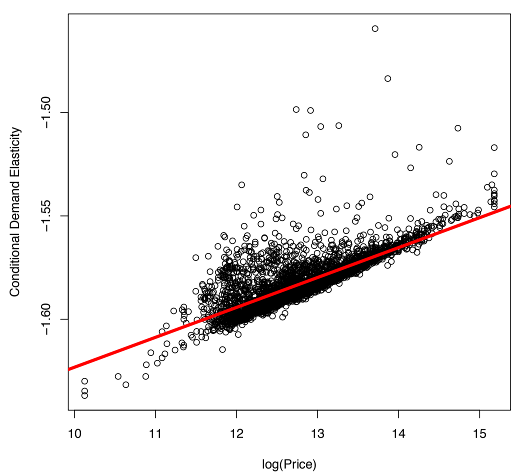
(b) Borrowing Elasticity



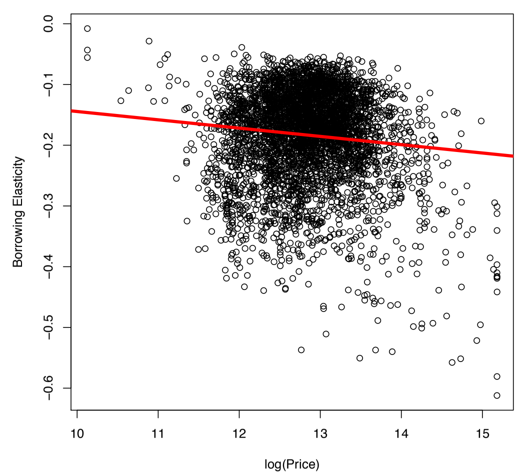
(c) Total Elasticity



(d) Conditional Elasticity vs. Price



(e) Borrowing Elasticity vs. Price



(f) Total Elasticity vs. Price

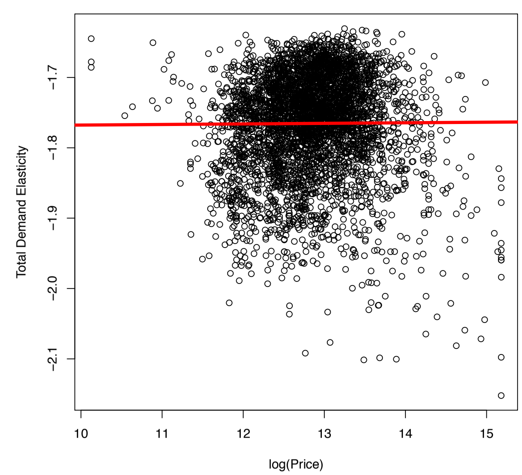
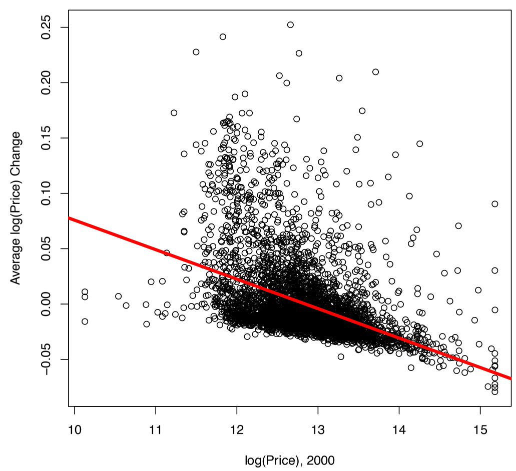


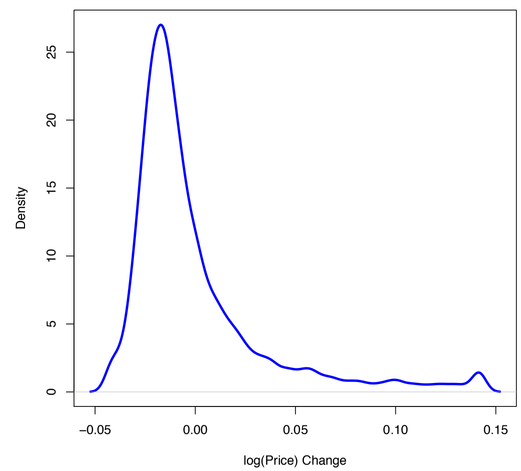
Figure 3: General Equilibrium Effects of Lending Standards on the Price Distribution

These figures illustrate the compression of the price distribution caused by the relaxation of lending standards. Figure (a) shows the relation between initial log(price) in 2000 and log(price) growth. Figure (b) shows the distribution of shifts $d\log(p)/d\psi$ across neighborhoods. Figure (c) shows the effect of lending standards on demand at constant prices, while Figure (d) shows the impact of lending standards on black demand for neighborhoods.

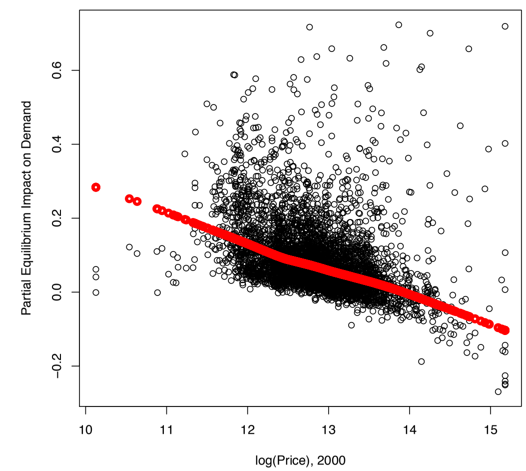
(a) G.E. Effect and Initial log(Price)



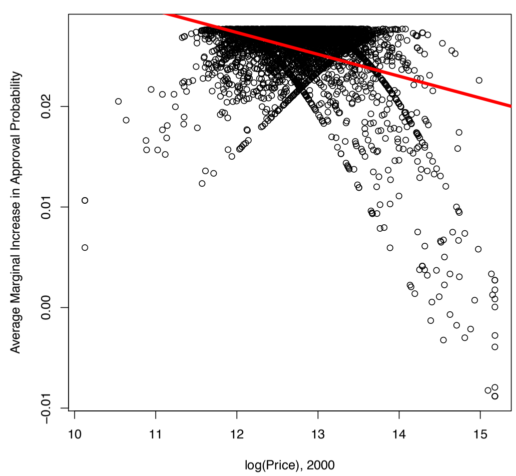
(b) Distribution of log(Price) Changes



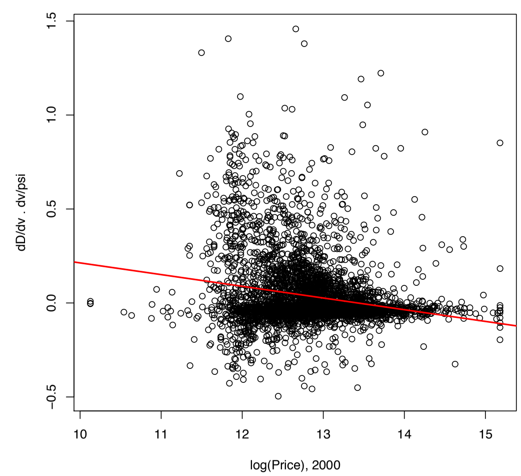
(c) Partial Equilibrium Demand Effects



(d) Approval Rate Increase and Initial log(Price)



(e) Social Interaction Impacts $\partial\mathbf{D}/\partial\mathbf{v}\cdot\partial\mathbf{v}/\partial\psi$



(f) Fraction Black and Initial log(Price)

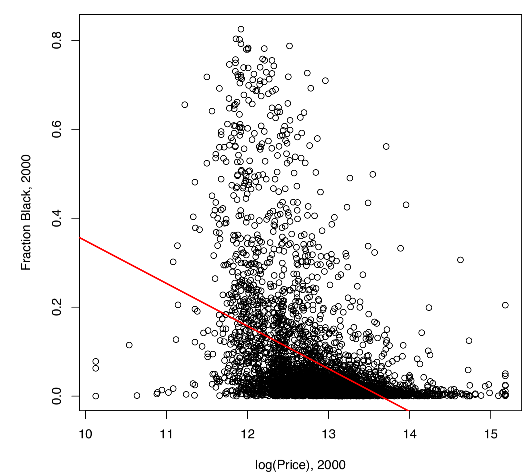
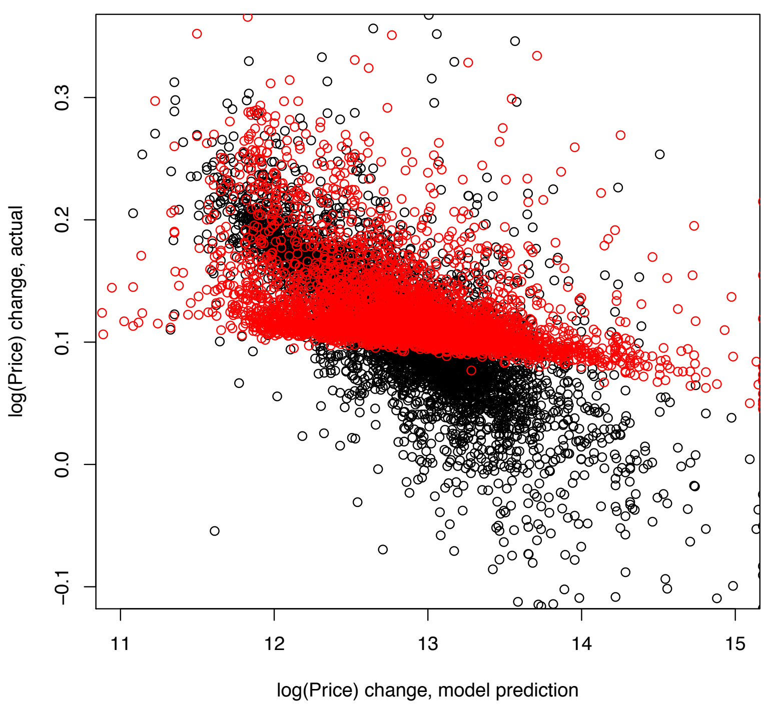


Figure 4: Compression of the Price Distribution – Comparison of General Equilibrium Predictions and the Actual 2000–2006 log(Price) Change

Black points below plot the actual *average annual change in the log(price) from our property-level transaction data against the initial log(price) in 2000. Red points plot the log(price) change as predicted by the model's general equilibrium comparative statics against the initial log(price) in 2000. Each point is a blockgroup.*

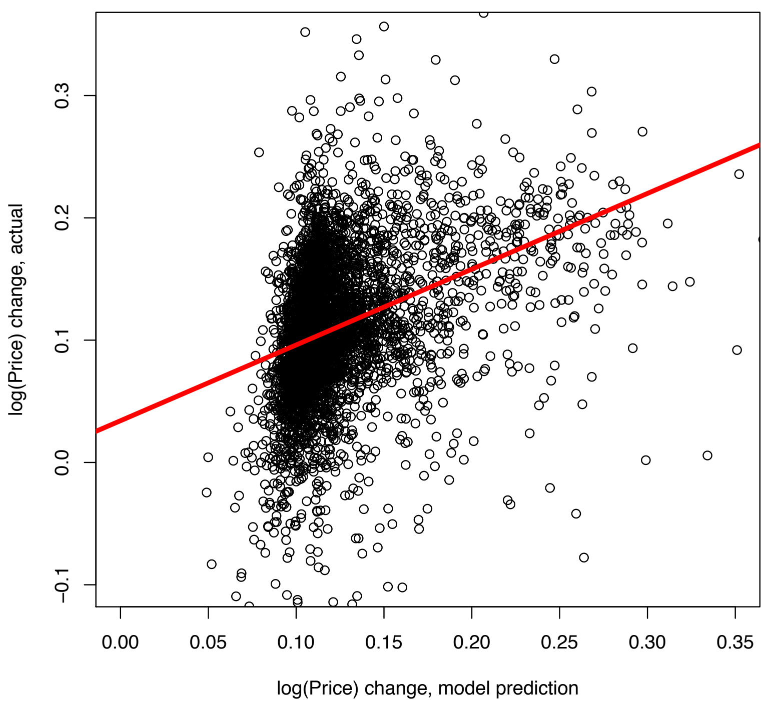


Actual log price change					
1st Qu.	Median	Mean	S.D.	3rd Qu.	Corr($\Delta log(p_{jt}), \log(p_{jt})$)
0.075	0.109	0.111	0.078	0.149	-0.630

Predicted log price change					
1st Qu.	Median	Mean	S.D.	3rd Qu.	Corr($\Delta log(p_{jt}), \log(p_{jt})$)
0.104	0.113	0.125	0.036	0.131	-0.435

Figure 5: General Equilibrium Predictions vs. Actual 2000–2006 log(Price) Change

The figure below plots, for each blockgroup, the realized average annual log(price) change against the predicted log(price) change. The red line is the corresponding linear regression.

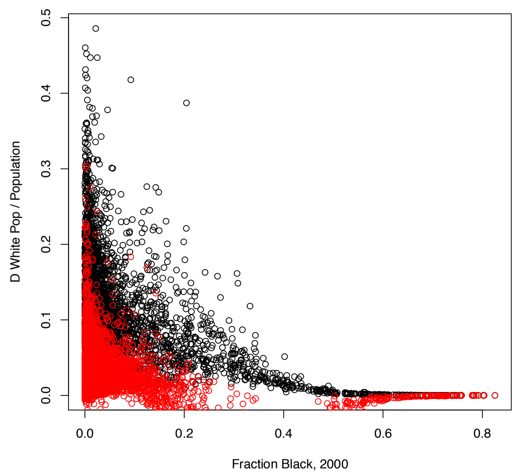


$$Actual\,\Delta\log(Price) = \underset{(0.004)}{0.034} + \underset{(0.031)}{0.619} \cdot Predicted\,\log(Price) + Residual, \quad R^2 = 0.083$$

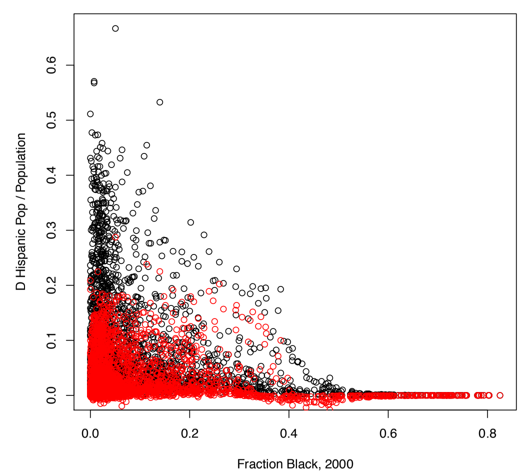
Figure 6: General Equilibrium Effects of Lending Standards on Segregation

These figures illustrate the impact of lending standards on segregation. The red dots are the demand changes accounting for the change in prices. The black dots are the demand changes at given prices.

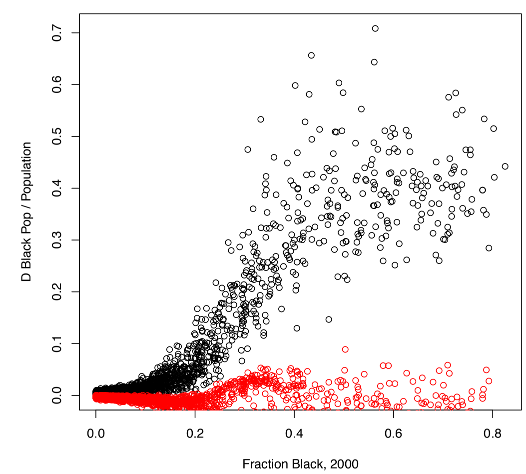
(a) White Demand Change and Initial Fraction Black



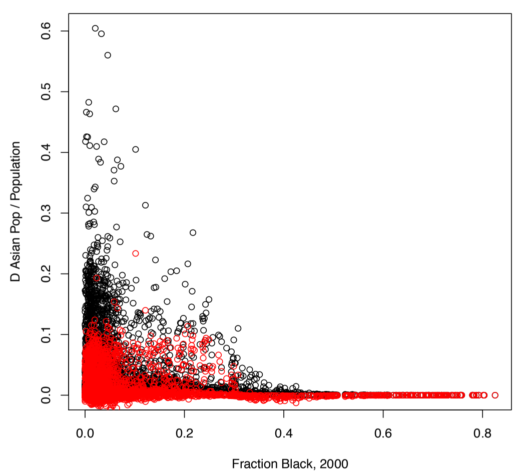
(b) Hispanic Demand Change and Initial Fraction Black



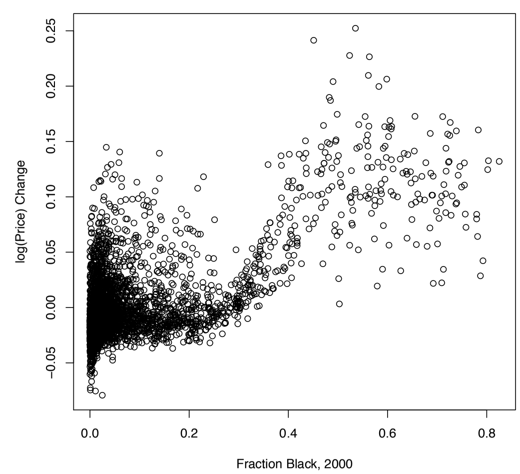
(c) Black Demand Change and Initial Fraction Black



(d) Asian Demand Change and Initial Fraction Black



(e) Price Change and Initial Fraction Black



(f) Hispanic Demand Change and Initial Fraction White

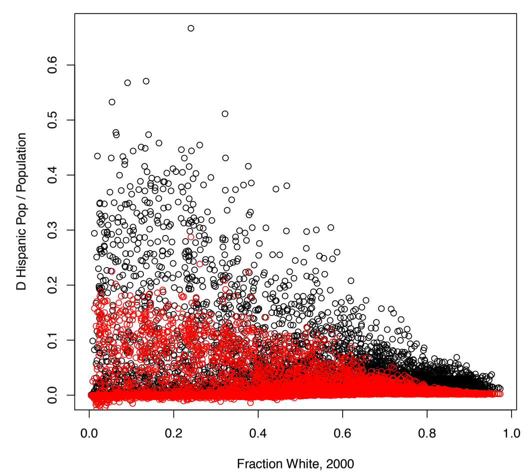


Table 1: Mortgage Approval Equation – Logit and IV Estimation – 1990–2010

The table presents estimation of the approval model (Specification 6). Top columns (1)–(2) (resp. (3)–(4)) present the specification without (resp., with) neighborhood-specific unobservable controls. The bottom part presents the IV regression using the LTI (resp., the liquidity) of the nearest branches in columns (1)–(2) (resp., (3)–(4)).

	Specification:			
	No Fixed Effect		Tract FE	
	Logit Coefficients	Marginal Probabilities	Logit Coefficients	Marginal Probabilities
LTI Ratio	-0.239***	-0.028***	-0.252***	-0.028***
	(0.008)	(0.001)	(0.008)	(0.001)
Black	-0.939***	-0.110***	-0.718***	-0.080***
	(0.031)	(0.004)	(0.031)	(0.003)
Asian	-0.121***	-0.014***	-0.066***	-0.007***
	(0.020)	(0.002)	(0.021)	(0.002)
Hispanic	-0.555***	-0.065***	-0.331***	-0.037***
	(0.025)	(0.003)	(0.029)	(0.003)
Other Race	-0.467***	-0.055***	-0.410***	-0.046***
	(0.025)	(0.003)	(0.024)	(0.003)
2000	-0.019	-0.002	0.026	0.003
	(0.023)	(0.003)	(0.024)	(0.003)
2010	0.005	0.001	0.031	0.003
	(0.027)	(0.003)	(0.031)	(0.003)
FHA Insured Loan	0.011	0.001	0.150***	0.017***
	(0.026)	(0.003)	(0.029)	(0.003)
Observations	208,936	208,936	208,206	208,206
Census Tracts	1,911	1,911	1,606	1,606
Pseudo R Squared	0.0186	0.0186	0.0436	0.0436

	IV Specification:			
	Loan-to-Income of Nearby Branches		Liquidity of Nearby Branches	
	Probit Coefficients	Marginal Probabilities	Probit Coefficients	Marginal Probabilities
Loan to Income Ratio	-1.007***	-0.179***	-0.791***	-0.132***
	(0.124)	0.014	(0.183)	(0.025)
Black	-0.331***	-0.059***	-0.461***	-0.077***
	(0.081)	(0.009)	(0.054)	(0.009)
Asian	0.116***	0.020***	-0.012	-0.002
	(0.039)	(0.008)	(0.044)	(0.007)
Hispanic	-0.053	-.009	-0.199***	-0.033***
	(0.051)	(0.008)	(0.046)	(0.007)
Other Race	-0.024	-0.004	-0.159***	-0.026***
	(0.057)	(0.009)	(0.058)	(0.009)
FHA Insured Loan	0.078	0.014	0.062	0.010
	(0.065)	(0.011)	(0.061)	(0.010)
VA-guaranteed	0.007	0.001	-0.082	-0.013
	(0.086)	(0.015)	(0.091)	(0.015)
FSA-RHS	0.597	0.106	0.554	0.092
	(0.383)	(0.067)	(0.440)	(0.073)
Year 2000	0.092	0.016	0.022	0.003
	(0.056)	(0.010)	(0.070)	(0.011)
Year 2010	0.628***	0.111	0.388**	0.065**
	(0.093)	(0.016)	(0.173)	(0.026)
Observations	41,153	41,153	41,153	41,153
Census Tracts	1,606	1,606	1,606	1,606
Wald χ^2	84.73	84.73	159.84	159.84

Table 2: Household Preferences and Willingness to Pay (WTP) for Local Amenities

The table reports estimates of the base utilities and their relationships with observable neighborhood amenities. We consider the model with borrowing constraints (column (1)), the model without borrowing constraints (column (2)). Column (4) reports the regression of the difference in base utilities in the two models.

	Dependent variable:			
	δ_{credit}	WTP, credit	$\delta_{no\ credit}$	WTP
Frac. Black	-11.928***	-7.53%	-10.235***	-5.88%
	(0.198)	-$ 32,871	(0.187)	-$ 25,647
× Black	15.758***	2.42%	13.754***	2.22%
	(0.089)	$ 10,555	(0.080)	$ 9,698
Frac. Hispanic	-1.993***	-1.26%	-2.066***	-1.19%
	(0.201)	-$ 5,492	(0.189)	-$ 5,177
× Hispanic	7.627***	3.56%	7.412***	3.38%
	(0.077)	$ 15,526	(0.073)	$ 14,732
Frac. Asian	-2.936***	-1.85%	-2.717***	-1.56%
	(0.172)	-$ 8,091	(0.162)	-$ 6,808
× Asian	7.643***	2.94%	7.295***	2.63%
	(0.087)	$ 12,971	(0.081)	$ 11,472
log(Median hh income)	0.636***	2.18%	0.622***	1.94%
	(0.046)	$ 9,499	(0.043)	$ 8,448
× (log(Individual Household Income) -	0.954***	-	0.934***	-
mean log(income))	(0.004)	-	(0.004)	-
Frac. College Graduate	0.562***	0.35%	0.529***	0.30%
	(0.121)	$ 1,549	(0.114)	$ 1,326
Median Age of Structure	-0.018***	-1.84%	-0.018***	-1.67%
	(0.001)	-$ 8,026	(0.001)	-$ 7,298
Median Number of Rooms	0.129	1.04%	0.149	1.09%
	(0.182)	$ 4,550	(0.172)	$ 4,779
Median Number of Rooms	-0.019	0.53%	-0.021	0.48%
× log(Distance to CBD)	(0.017)	$ 2,306	(0.016)	$ 2,111
School Performance	0.061***	0.75%	0.025***	0.28%
(*Academic Performance Index*)	(0.009)	$ 3,265	(0.001)	$ 1,217
× (log(Individual Household Income) -	-0.227***	-	-0.1544***	-
mean log(income))	(0.003)	-	(0.003)	-
log(Median price)	-1.584***	-	-1.742***	-
	(0.101)		(0.095)	
σ(log(Median price))	0.138***	-	0.113***	-
	(0.025)		(0.028)	
Observations	13,154	-	13,154	-
Neighborhoods (Blockgroups)	4,416	-	4,416	-
Years	3	-	3	-
R Squared	0.336	-	0.293	-

Standard errors clustered by block group in parenthesis.
***: Significant at 1%, **: Significant at 5%, *: Significant at 10%.

Table 3: General Equilibrium Effects of Lending Standards on Racial Segregation

The table presents the comparative statics impact of the marginal change in lending standards on racial segregation. Table (a) is the impact on racial exposure at given prices, and Table (b) is the total impact, accounting for price changes, on racial exposure (Section (5.2)). The tables read as follows: the cell in the second row and the first column is the impact of lending standards on the exposure of Blacks to Whites, in percentage points. The diagonal of each table is the impact on isolation, i.e. the exposure of a racial group to same-race neighbors.

(a) Partial Equilibrium Segregation Changes

	Exposure to			
Exposure of	White	Black	Hispanic	Asian
White	+12.77 ppt	+1.47 ppt	+4.71 ppt	+3.52 ppt
Black	+5.72 ppt	+13.92 ppt	+4.40 ppt	+2.70 ppt
Hispanic	+6.86 ppt	+2.51 ppt	+10.74 ppt	+2.78 ppt
Asian	+7.77 ppt	+1.83 ppt	+4.12 ppt	+8.61 ppt

(b) General Equilibrium Segregation Changes

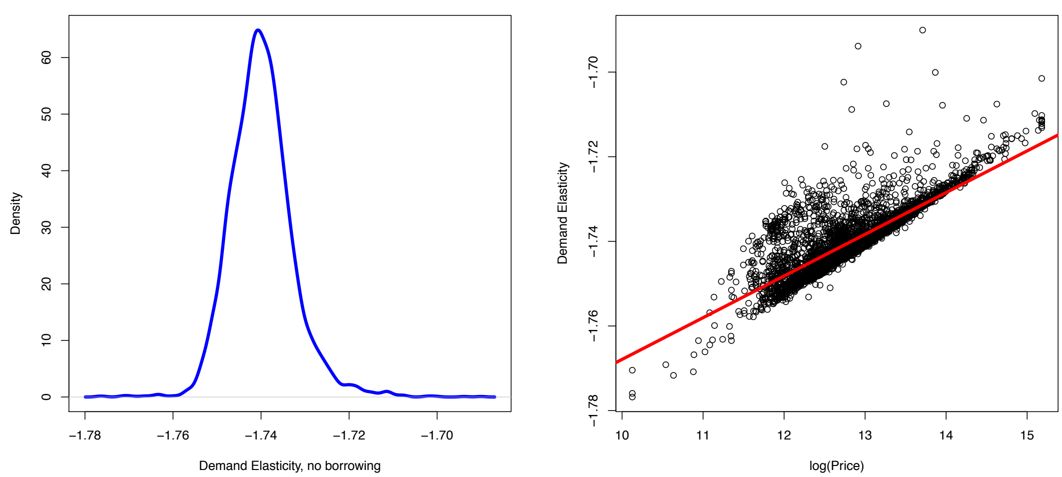
	Exposure to			
Exposure of	White	Black	Hispanic	Asian
White	–2.02 ppt	–0.36 ppt	+1.65 ppt	+0.71 ppt
Black	–2.04 ppt	–0.62 ppt	+1.90 ppt	+0.76 ppt
Hispanic	+1.42 ppt	–0.38 ppt	–1.56 ppt	+0.52 ppt
Asian	+1.80 ppt	–0.32 ppt	+1.42 ppt	–2.90 ppt

Appendix Figure A: Demand Elasticity Distributions – Model without Borrowing Constraints

This Figure presents the elasticity of demand (w.r.t. own price) in the model without borrowing constraints. This amounts to constraining the probability of approval to 1 for every neighborhood and for every household. Figure (a) is the distribution of neighborhood demand elasticities in 2000. In Figure (b) each dot is a neighborhood in 2000, and the red line is the OLS regression of demand elasticity on current log(price) in 2000.

(a) Demand Elasticity

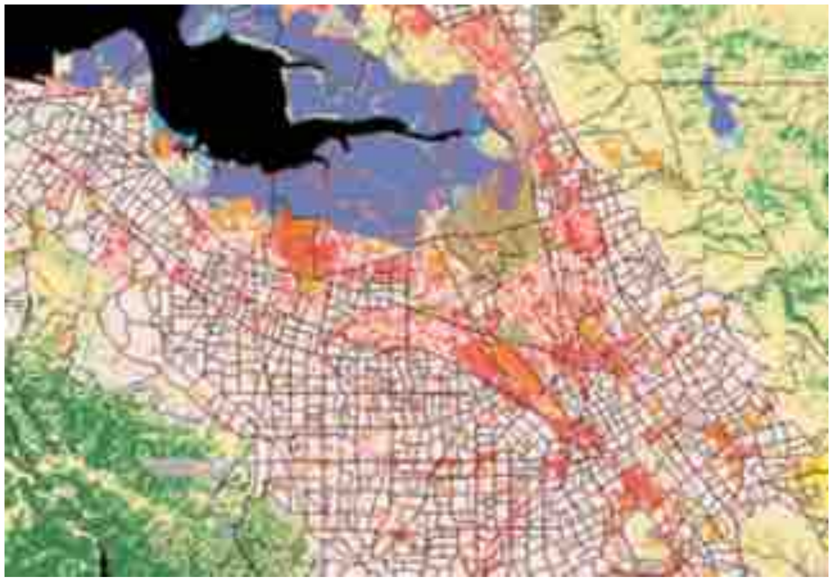
(b) Demand Elasticity vs. log(Price)



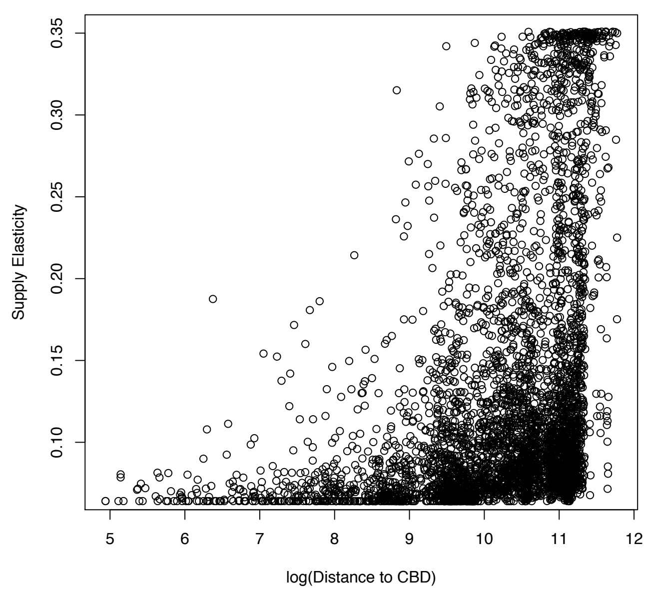
Appendix Figure B: Undeveloped Land - San Francisco Bay Area

The top figure presents a scatter plot of the log(distance to the Central Business District) against neighborhood-level housing supply elasticity. Each circle is a neighborhood. The bottom figure presents the source satellite data (landcover USGS) used to measure the undeveloped share of blockgroups; zoomed on San Jose. Each pixel is a 30m×30m cell. The color coding is as follows: Light Red = Low Intensity Residential Development Dark Red = High Intensity Residential Development Orange = Commercial or Industrial Development Light Blue = Water Dark Blue = Barren Land.

(a) Undeveloped Land – Focus on the San Jose Metropolitan Area



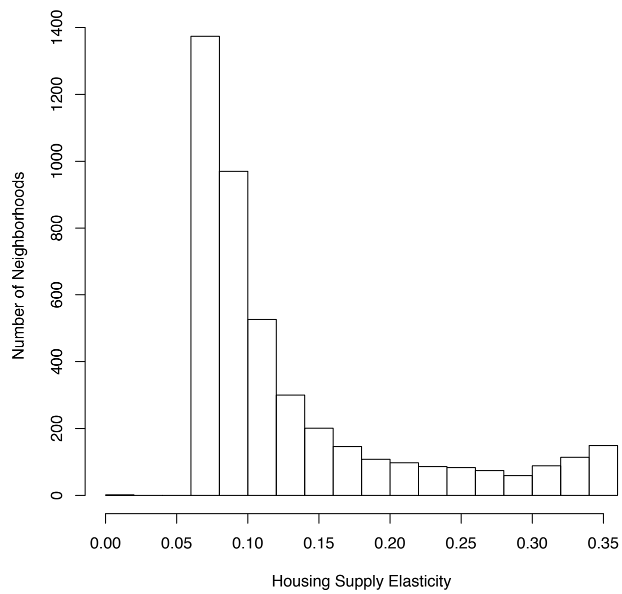
(b) Distance to Central Business District and Elasticity



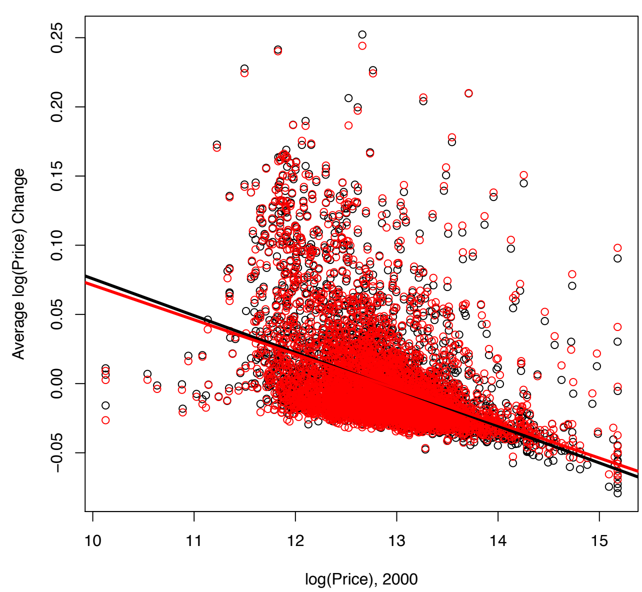
Appendix Figure C: General Equilibrium Price Changes with Neighborhood-Level Housing Supply Elasticity

The top figure presents estimates of blockgroup- (neighborhood-) specific housing supply elasticity measures. The estimation of such elasticities, based on undeveloped land data from USGS, is described in Section 6.1. The bottom figure presents general equilibrium effects that account for local housing supply elasticities.

(a) Distribution of Estimated Local Supply Elasticities



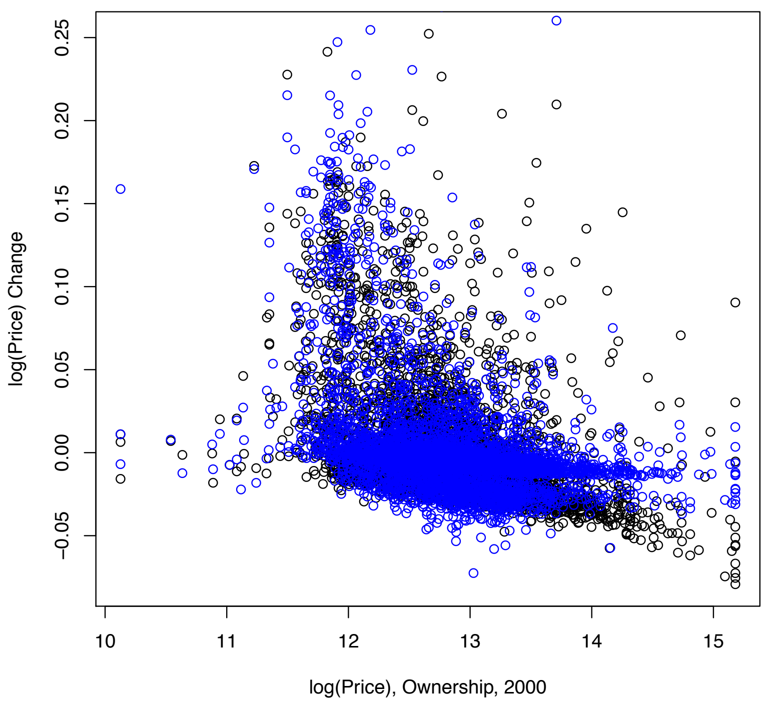
(b) Impact of Housing Supply Elasticities on Estimated General Equilibrium Effects



Appendix Figure D: Tenure Choice and The Relaxation of Lending Standards

The Figure shows the general equilibrium impact of the relaxation of lending standards on prices, with the option of renting in each neighborhood. A relaxation of lending standards increases households' access to homeownership in partial equilibrium, but leads to price increases for owner-occupied units.

Blue points are the general equilibrium log(price) changes in the model with both tenure and neighborhood choice, and credit constraints for access to homeownership. Black points as in the baseline model Figure 3a.



Appendix Table A: Summary Statistics and Data Sources

The first part of this table presents neighborhood data from a variety of sources: Summary File 3 of the 1990 and 2000 Censuses and employment numbers from the ZIP-level County Business Patterns. The second part is household data from the 1% Census microfiles in 1990, 2000, 2010 for the San Francisco MSA. The third part presents mortgage application data from the Home Mortgage Disclosure Act of 1990 and 2000.

	Mean	Median	S.D.	Min	Max	Obs.
Blockgroup Data						
log(Median price)	12.757	12.737	0.688	7.601	18.258	13,154
Median number of rooms	5.156	5.062	1.254	0.074	9.100	13,154
Median age of structure	33.562	35.000	17.052	0.000	61.000	13,154
Frac. Black	0.078	0.025	0.140	0.000	0.948	13,154
Frac. Hispanic	0.278	0.170	0.289	0.002	1.818	13,154
Frac. Asian	0.175	0.116	0.163	0.000	0.983	13,154
log(Median household income)	11.050	11.053	0.542	6.787	12.429	13,154
Frac. college educated	0.439	0.426	0.211	0.000	0.972	13,154
Frac. more than high school	0.582	0.517	0.251	0.033	1.000	13,154
Frac. denied	0.132	0.141	0.052	0.000	0.615	13,154
Micro Census Data						
Household Income	168,300	68,440	864,759	10,000	100,000	120,029
White	0.555	1.000	0.497	0.000	1.000	120,029
Black	0.062	0.000	0.240	0.000	1.000	120,029
Hispanic	0.157	0.000	0.364	0.000	1.000	120,029
Asian	0.201	0.000	0.401	0.000	1.000	120,029
Mortgage Application Data						
Approved	0.769	1.000	0.421	0.000	1.000	163,630
Loan to Income Ratio	3.380	3.267	0.874	1.883	6.117	163,630
Loan Amount ('000)	424.503	436.000	132.960	78.000	638.000	163,630
Applicant Income ('000)	162.710	150.000	73.531	44.000	406.000	163,630
Black	0.063	0.000	0.243	0.000	1.000	163,630
Hispanic	0.021	0.000	0.143	0.000	1.000	163,630
Asian	0.209	0.000	0.406	0.000	1.000	163,630
FHA Loan	0.001	0.000	0.018	0.000	1.000	163,630

Appendix Table B: Mortgage Approval Equation – Probit Regressions – 1990–2010

The table presents probit estimation of the approval model (Specification 6). Columns (1)–(2) present the coefficients and marginal probabilities when using the full sample. Columns (3)–(4) present results when using the sample of applications to banks regulated by the Federal Reserve, the Office of the Comptroller of the Currency, and the Federal Deposit Insurance Corporation, and with census tract information.

	Specification:			
	Probit Tract F.e.		Probit Tract F.e. on Subsample	
	Probit Coefficients	Marginal Probabilities	Probit Coefficients Coefficients	Marginal Probabilities
Loan to Income Ratio	-0.136***	-0.029***	-0.156***	-0.021***
	(0.004)	(0.001)	(0.009)	(0.001)
Black	-0.403***	-0.086***	-0.384***	-0.051***
	(0.018)	(0.004)	(0.050)	(0.006)
Asian	-0.037***	-0.008***	-0.066**	-0.008**
	(0.011)	(0.002)	(0.030)	(0.004)
Hispanic	-0.184***	-0.039***	-0.169***	-0.022***
	(0.016)	(0.003)	(0.042)	(0.005)
Other Race	-0.224***	-0.048***	-0.248***	-0.033***
	(0.013)	(0.003)	(0.038)	(0.005)
FHA Insured Loan	0.077***	0.016***	0.199***	0.027***
	(0.016)	(0.003)	(0.064)	(0.008)
VA-guaranteed	0.098***	0.021***	-0.075	-0.010
	(0.038)	(0.008)	(0.092)	(0.012)
FSA-RHS	-0.242	-0.051	0.746	-0.010
	(0.168)	(0.036)	(0.556)	(0.075)
Year 2000	0.020	0.004	-0.082	-0.011
	(0.013)	(0.003)	(0.120)	(0.016)
Year 2010	0.020	0.004	-0.025	-0.003
	(0.016)	(0.003)	(0.131)	(0.017)
Observations	208,206	208,206	41,153	41,153
Census Tracts	1,606	1,606	1,603	1,603
Pseudo R Squared	0.0434	0.0434	0.1762	0.1762

Appendix Table C: Mortgage Approval Equation – Additional Instrumental Variable Strategies – 1990–2010

The table presents instrumental variable estimation of the approval model (Specification 6). Columns (1)–(2) present the coefficients and marginal probabilities when using the liquidity of nearby bank branches that were established prior to 1985. Columns (3)–(4) present results when using the racial composition of nearby tracts as an instrument for the racial dummies of the applicant. The Wald χ^2 tests the null hypothesis that all coefficients are equal to zero.

	IV Specification: Liquidity of Nearby Branches Set Up before 1985		Branches IV and Racial Composition of Adjacent Neighborhoods	
	Probit Coefficients	Marginal Probabilities	Probit Coefficients	Marginal Probabilities
Loan to Income Ratio	-1.049***	-0.225***	-1.037***	-0.228***
	(0.209)	(0.042)	(0.205)	(0.044)
Black	-0.364***	-0.078***	-1.247***	-0.274***
	(0.056)	(0.011)	(0.266)	(0.056)
Asian	0.038	0.008	0.320**	0.070**
	(0.044)	(0.009)	(0.153)	(0.035)
Hispanic	-0.121**	-0.026**	-0.713***	-0.156***
	(0.049)	(0.010)	(0.222)	(0.046)
Other Race	-0.064	-0.013	-0.080	-0.017
	(0.063)	(0.013)	(0.086)	(0.018)
FHA Insured Loan	0.004	0.001	0.170***	0.037***
	(0.057)	(0.012)	(0.060)	(0.013)
VA-guaranteed	-0.025	-0.005	0.097	0.021
	(0.081)	(0.017)	(0.098)	(0.022)
FSA-RHS	0.446	0.095	0.466	0.102
	(0.318)	(0.067)	(0.297)	(0.064)
Year 2000	0.075	0.016	0.118***	0.026***
	(0.051)	(0.011)	(0.038)	(0.008)
Year 2010	0.686	0.147***	0.545***	0.119***
	(0.203)	(0.042)	(0.193)	(0.042)
Observations	41,153	41,153	41,153	41,153
Census Tracts	1,606	1,606	1,606	1,606
Wald χ^2	150.25	150.25	152.37	152.37

Appendix Table D: Change in Lending Standards between 2000 and 2006

The table presents estimation of the approval model (Specification 6), with the 2006 mortgage application data.

	Approved	Change 2000 – 2006
log(Price)	–0.352*	+0.159
	(0.127)	(0.140)
log(Income)	0.759*	–0.503*
	(0.110)	(0.115)
Constant	–2.111	+2.967*
	(0.962)	(1.113)
Black or African American, Nonhispanic	-1.154*	+0.345*
	(0.086)	(0.093)
Hispanic, Any Race	–0.716*	+0.294**
	(0.103)	(0.117)
Asian, Nonhispanic	–0.286*	+0.404*
	(0.069)	(0.075)
Observations	103,176	
Census Tracts	1,315	
Pseudo R Squared	0.0407	

A Appendix: Equilibrium Properties

Proposition 1. ***(Equilibrium Existence)*** *There exists an equilibrium vector of prices and neighborhood demographics* $(\mathbf{p}^* = (p^*_{jt}), \mathbf{v}^* = (\mathbf{v}^*_{jt}))_{j=1,2,\ldots,J,\,t=1,2,\ldots,T}$ *that satisfies conditions (13) and (14).*

Proof. For notational convenience, and without loss of generality, we consider a scalar rather than vector v_j, i.e. households have a preference for neighbors of only one demographic subgroup. We start by proving existence with an non-perfectly elastic housing supply, and then turn to equilibrium existence with a perfectly elastic housing supply. First, note that the equilibrium of the city can be equivalently rewritten as:

$$
\begin{aligned}
s_1^{-1}(D_1(p_1, \ldots, p_J, v_1, \ldots, v_J) &= p_1 \\
&\vdots \\
s_J^{-1}(D_J(p_1, \ldots, p_J, v_1, \ldots, v_J)) &= p_J \\
D_1^W(p_1, \ldots, p_J, v_1, \ldots, v_J) &= W_1 s_1 \\
&\vdots \\
D_2^W(p_1, \ldots, p_J, v_1, \ldots, v_J) &= W_J s_J
\end{aligned}
$$

and define the mapping: $\phi : \mathbb{R}^J \times [0,1]^J \to \mathbb{R}^J \times [0,1]^J$. The mapping is continuous given the functional forms of the supply curves, the demand curves, and the probability of origination. Notice that the upper bound of D_j is 1 for each function, hence the mapping ϕ takes its values in $[s^{-1}(0); s^{-1}(1)]^J \times [0,1]^J$, which is a closed compact subset of $\mathbb{R}^J \times [0,1]^J$. We can thus consider the mapping $\tilde{\phi}$ of $[s^{-1}(0); s^{-1}(1)]^J \times [0,1]^J$ to itself, equal to ϕ on $[s^{-1}(0); s^{-1}(1)]^J \times [0,1]^J$. Such a mapping is continuous, and $[s^{-1}(0); s^{-1}(1)]^J \times [0,1]^J$ is a convex set of a Banach space. Hence, by the Brouwer fixed point theorem, $\tilde{\phi}$ admits a fixed point, i.e. a vector $(p_1^*, \ldots, p_J^*, v_1^*, \ldots, v_J^*)$ that satisfies the $2J$ equations that define the equilibrium.

We then consider the case of a perfectly inelastic housing supply. Set $s_j(p) = s_j \cdot p^{\eta^s}$ with η^s the elasticity of housing supply. For a given η^s consider the set of equilibrium vectors $E(\eta^s) = \{(p_1^*, \ldots, p_J^*, v_1^*, \ldots, v_J^*)\}$. We just showed that $E(\eta^s) \neq \emptyset$ for any $\eta^s > 0$. Consider a sequence of equilibrium vectors for a sequence of $\eta^s \to 0$. Such sequence of equilibrium vectors converges to a vector $(\mathbf{p}^* = (p^*_{jt}), \mathbf{v}^* = (\mathbf{v}^*_{jt}))_{j=1,2,\ldots,J,\,t=1,2,\ldots,T}$ that is an equilibrium price vector when $\eta^s = 0$. $\square$

Proposition 2. ***(Global Equilibrium Uniqueness with no Social Preferences)*** *Where there are no preferences for same-race neighbors ($\gamma_i = 0$), the city equilibrium is unique, up to one price.*

Proof. To prove such equilibrium uniqueness, we extend the model by treating consumer income as a household endowment ω_i for each household i. Consumer demand for neighborhood j is $D_j(p_1, \ldots, p_J, p)$, where p is the price of the numeraire consumer good. Thus the value of household i's endowment is $p\omega_i$ in terms of the numeraire good. Now notice that demand for neighborhood j is homogeneous, as the approval specification depends on the ratio of price and income and neighborhood demands are left unchanged when all prices are multiplied by a constant. Thus p can be normalized to 1. Note also that as the price of neighborhood j increases, demand for neighborhood j strictly decreases (the probability of acceptance goes down strictly and the utility value of neighborhood j strictly decreases). When the price of neighborhood $-j$ strictly increases, demand for neighborhood j strictly increases (the probability of acceptance in neighborhood $-j$ goes down strictly, and the utility value of neighborhood $-j$ strictly decreases). Thus housing in neighborhood 1 and housing in neighborhood 2 are gross substitutes. By Proposition 17.F.3 of Mas-Colell, Whinston, Green et al. (1995), the equilibrium of the city is unique up to one neighborhood price. $\square$

B Appendix: Estimation Technique

We estimate the model using the nested fixed point (NFP) method of Berry et al. (1995). Su & Judd (2012) has shown that the NFP and the MPEC method of Dubé, Fox & Su (2012) are equivalent, but the NFP method requires a tight tolerance level to converge. The inner loop of our contraction mapping uses a 10^{-12} tolerance, and the outer loop a 10^{-6} tolerance level.

Estimating the model in double precision Fortran has considerable impacts on speed and precision. The algorithm is available through the corresponding author. The contraction mapping is also parallelized by acknowledging that the demand for housing is independent across the three decades of the data, 1990, 2000, and 2010. For each computation of $\boldsymbol{\delta} = (\boldsymbol{\delta}_{1990}, \boldsymbol{\delta}_{2000}, \boldsymbol{\delta}_{2010})$ therefore, the R package `snowfall` parallelizes the three contraction mappings.

Minimization of the objective function $G(\boldsymbol{\theta})$ proceeds using the Nelder-Mead algorithm, and the nature of the optimum $\hat{\boldsymbol{\theta}}$ (global vs. local) is assessed using profile GMM objective functions.

C General Equilibrium Derivations

This Appendix's section provides the derivations for Section 5.1(ii): the impact of a log(price) change on demand, at given lending standards. Equation (25) has two parts: the shift in the choice set probabilities implied by the shift in price (without a change in lending standards), and the shift in demand caused by a shift in utility at given choice set probabilities. We detail both below.

Shift in choice set induced by changing prices.

We first focus on the first term, which is the impact of the log price $log(p_k)$ on choice set probabilities. For every choice set that includes neighborhood k the probability of that choice set goes down, and for every choice set that does *not* include neighborhood k, the probability of that choice set goes up.

$$\frac{\partial}{\partial \log(p_k)} P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi) \quad = \quad -\alpha_{approval}(\mathbf{1}(k \in C) - \phi_{kt}) P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi)$$

Hence the impact of a price change on demand at given utility levels. When estimating the impact of prices on demand at given lending standards, this is the first term (*"shift in choice set"*) in equation 25:

$$\begin{aligned}
&\frac{\partial}{\partial \log(p_k)}\ D(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}; \psi)|_{fixed\,utility} \\
= &\sum_{C \in \mathbb{C}} \int_X \alpha_{approval}\left(\mathbf{1}(k \in C) - \phi_{kt}\right) P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi) \cdot P(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C_{it}) f(\mathbf{x}_{it}) d\mathbf{x}_{it}
\end{aligned}$$

Effect of prices on neighborhood choice within the choice set.

We then turn to the second term, i.e. the impact of prices on utility, thus to the impact of prices on households' choice *within their choice set.* The derivative of conditional demand $P(j,t)$ w.r.t. its own price p_j is $\alpha_{utility} \cdot P(j,t) \cdot (1 - P(j,t))$. The coefficient α is the coefficient of log price in utility. Thus the second term ("*shift in utility*") in equation 25:

$$\frac{\partial}{\partial \log(p_j)} D(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}; \psi) \quad = \quad -\sum_{C \in \mathbf{C}} \int_X \alpha(\mathbf{x}_{it}, \tilde{\boldsymbol{\beta}}_{it}) \cdot P(C|\mathbf{z}_{\cdot t}, \mathbf{x}_{it}; \psi)$$
$$\cdot P(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C_{it}) \cdot (1 - P(j,t|\boldsymbol{\delta}_{\cdot t}, \mathbf{z}_{\cdot t}, \mathbf{x}_{it}, C_{it})) f(\mathbf{x}_{it}) d\mathbf{x}_{it}$$

We apply this same logic to calculate the impact of the price of another neighborhood on its own demand, and thus obtain cross-price elasticities.